omega

The International Journal of Management Science

Volume 7 1979

Author and Title Index

● **Chief Editor**
Professor SAMUEL EILON, Department of Management Science,
DSc(Eng.), PhD Imperial College, London, SW7 2BX

● **Associate Editors**
RJ BETTS and **RB FLAVELL** Department of Management Science,
Imperial College, London, SW7 2BX

Editorial Advisory Board

Professor RJ BALL
London Graduate School of Business Studies, Sussex Place, Regent's Park, London NW1 4SA

Professor BEN BERNHOLTZ
Department of Industrial Engineering, University of Toronto, Toronto 5

Professor C WEST CHURCHMAN
Space Sciences Laboratory, University of California, Berkeley, CA 94720

Professor WW COOPER
Graduate School of Business Administration, Harvard University, Boston, MA 02163

Professor BELA GOLD
Department of Economics, Case Western Reserve University, Cleveland, OH 44106

Professor CH GUDNASON
Danmarks Tekniske Højskole, København

Professor DOUGLAS C HAGUE
Manchester Business School, Booth Street West, Manchester M15 6PB

DAVID B HERTZ
McKinsey and Company Inc., 245 Park Avenue, New York, NY 10017

Professor KNUT HOLT
The Norwegian Institute of Technology, 7034 Trondheim

Professor HAROLD J LEAVITT
Graduate School of Business, Stanford University, Stanford, CA 94305

Professor GERALD J LIEBERMAN
Department of Operations Research, Stanford University, Stanford CA 94305

Professor RINTARO MURAMATSU
Department of Industrial Engineering, Waseda University, 4-170 Nishiokubo, Sinjuku-ku, Tokyo 160

NAOMI YAMAKI, President
Mitsubishi Space Software Co Ltd, 2-4-1, Hamamatsu-Cho, Minato-Ku, Tokyo 105

PERGAMON PRESS
Oxford · New York · Toronto · Paris
Frankfurt · Sydney

AUTHOR INDEX

Ackoff RL 179
Amara R 273
Arcade SH 355
Atherton J 104

Babu S 153
Barbaroussis N 82
Barrett M 431
Baum S 249
Bechtold SE 565
Beckett T 101
Bennett PG 309
Bhandari L 359
Bowen KC 489, 503
Bradley SP 77
Brown DJ 55
Bryer RA 219
Burton RM 113
Button KJ 399

Carlson R 249
Chiang DT 287
Common MS 385
Corner LJ 257
Cosmetatos G 82

Damon WW 113
Davies D 275
Doholakia N 359
Duncan CP 345

Eden C 119
Eilon S 1, 82, 89, 166, 173, 269, 481
Ellis PM 573

Finney MJ 333
Firth M 129
Flavell R 25
Foy N 180
Frey SC 77

Gentry JW 241
Giesen MO 309
Ginzberg MJ 527
Glisson CA 180
Gordon LA 67
Graham RJ 61
Gray A 451
Greer WR Jr 361
Gustavsson SO 179

Haessler RW 145
Heald D 469
Hinomoto H 567

Jackson D 43
Jones WH 9

Keeney RL 191, 364
Keown AJ 345
Khurana R 359
King JR 233
Kleijnen JPC 539
Kottas JF 80
Kulkarni RB 191
Kwak NK 367

LaCava GJ 361
Larcker DF 67
Larrèché J-C 207
Lau H-S 80
Levine G 321
Lewis D 421
Lowe J 421

Mann JR 513
Meade N 333
Mitchell GH 513
Moore PG 183
Morse JJ 489
Moseley R 105

Nahmias S 321
Nair K 191
Narula S 137
Nash CA 441

Oakley MH 9
Ogbu U 137

Pearce DW 379
Penn E 25

Rivett BHP 33

Salkin GR 25
Schniederjans MJ 367
Sculli D 557
Seltzer J 61
Shanteau J 241
Shirland LE 573
Simpson DRF 371
Sims D 119
Soenen LA 339
Starr MK 490
Steele R 451
Steger WA 561
Stratton A 565
Strong EC 207
Sweet T 43

Tadikamalla PR 560
Taylor BW III 345
Treddenick JM 459
Tuggle FD 67
Turner RK 411

Vogel MA 559
Vora MN 359
Vrat P 153

Walbank M 297
Wensley JRC 15
Williams HP 163
Wills G 104
Woolley RN 574

TITLE INDEX

Ackoff's Fables (Editorial) 89
Ackoff's Fables (Feedback) 179
Ackoff's Fables (Feedback) 490
Allocation of Research and Development Funds: a Zero–One Goal Programming Approach 345
Analysis of the Choice between Domestic and Foreign Loan Facilities 333
Application of Catastrophe Theory to Management Science Process 61
Aristotle's Fallacy: a Hypergame in the Oil Shipping Business 309
Assessment of Fifteen Years of Urban Modeling 545

Budgets for Integrating Decentralized Investment and Production Planning 113

Capital Rationing and Discounting in Public Sector Investment Decisions 77
Capital Rationing Decision with Internal and External Funds 567
Case Study in Venture Management 9
Change, Clarity and Clausewitz 275
Cost–Benefit Analysis—a Critique 411
Cost–Effectiveness in Defence Expenditure 459
Creativity and Operational Research 574

Decision Model with Stochastic Cost and Demand Curves 80
Disclosure of Information by Companies 129
Dynamic Aspects of Line Balancing 557

Economic Approaches to Environmental Decision Making 421
Efficient Use of Non-Renewable Resources 371
Energy Forecasting 493
Environmental Impact Statements 431
Evaluating Management Education (Editorial) 481
Evaluation of Management Information Systems 557

Exercise in Probability Encoding 355
External Costs: Myth or Rationale for State Intervention? 385

Four Examples of Artless Modelling 163
Futures Literature and Futures Forecasting—a Critical Review 33

Goal Programming Model for Improved Transportation Problem Solutions 367

Handling Financial Risks 183
Hierarchal Location–Allocation Problem 137

Improving MIS Project Selection 527
In Defense of Determinism (Feedback) 180
Increasing the Effectiveness of Modeling 521
Industrial Relations Problems—OR's Contribution 513
Information as a Tangible Benefit (Feedback) 104
Informational Impediments to the Use of Capital Budgeting Models 67
Integrating Model of Motivated Work Behaviour 297
Inventory Control of Spares in a Transport Corporation—a Case Study 153
Inventory Management and Cost-Benefit Analysis: a Case Study 321

Lateral Consistency? (Editorial) 173
Linear Programming: Some Unsuccessful Applications 257
Liquidity Planning in a Small Bank 287
Lognormal Approximation to the Lead Time Demand in Inventory Control 553

Modelling the Use of the Manager's Time 55
Models for Decision-Making in the Public Sector 399

Normal Approximations for the Greater of Two Normal Random Variables 361

On the Nature of Problems in Consulting Practice 119
On Normative Policy Models in Developing Countries 359
On the Optimum Location of Training Centers: a Case Study 569
On Solutions that are Better Than Most 249

People and Size (Feedback) 101
Personal and Organizational Value Systems: How Should we Treat These in OR Studies? 503
Planning, Futures, and the Skeptics (Feedback) 273
Planning of a Department Store 25
Political Man (Feedback) 180
Portfolio Model for Foreign Exchange Exposure Management 339
Power Corrupts (Editorial) 269
Programme Budgeting in the Health Sector 451
Public Expenditure and Public Accountability 469
Public Risk from Potential LNG Spills 364
Public Sector Decision-Making (Editorial) 379

Quantitative Models for Optimal Rest Period Scheduling: a Note 565
Question of Size (Editorial) 1
Question of Size (Feedback) 179
Question of Size (Feedback) 489

Relative Success of Marketing Decision Aids: a Survey 15
Revision of Consumer Intention: Two Models 241
Risk Analysis of an LNG Terminal 191

Scheduling and the Problem of Computational Complexity 233
Solving the Two-Stage Cutting Stock Problem 145
Some Useful Ratios in Evaluating Performance 166
State of Variable Costs in Profitability Planning 82
Status of the Systems Approach 219
Supportive Application of Management Science to Marketing Strategy Formulation 207

Technical Change and Employment in the Post-War Gas Industry 105

Use of Economic Appraisal in Transportation Analysis 441

Vroom–Yetton Leadership Model (Feedback) 489

World Strike Wave 43

Aims and Scope

The purpose of OMEGA is to provide a vehicle for reporting developments in management science, operational research and managerial economics, including research results and applications. Published material ranges from original contributions to review articles describing the state of the art in specific areas, together with shorter critical assessments of particular management techniques. Other features of the journal are the Memoranda section for short communications, Feedback (a correspondence column) and a register for theses and dissertations.

The journal is intended to serve the needs of practising managers, specialists in management services, operational research workers and management scientists, management consultants, academics, students and research personnel.

Contributions may be made in any of the following categories:

1. *State of the art.* These papers would normally be more than 5000 words long, devoted to an interpretation and evaluation of existing knowledge in a particular field or function of management. Such papers should aim to be valuable both to practising managers and also to research workers and should therefore include a detailed bibliography.

2. *Theory.* This category caters for papers in specialised fields, 1500–5000 words long, which report research results and new theoretical developments.

3. *Applications.* These papers describe interesting applications of management science techniques and should typically be 1500–5000 words in length.

4. *Memoranda.* This section is devoted to short contributions, preferably under 600 words, to preliminary communications for new ideas and propositions, technical notes, brief descriptions of internal research reports or theses/dissertations, extension or critique of widely accepted concepts, brief accounts of applications and implementation. The intention is to publish such memoranda with as little delay as possible. Items for Memoranda Section should have the minimum number of diagrams (preferably none if possible) and lists of references should also be pared to the absolute minimum.

5. *Feedback.* Correspondence is invited on issues of significance in the field of management. This need not necessarily relate to material previously published in OMEGA.

6. *International Register of Theses and Dissertations.* Information for inclusion in this Section should be presented in the style of the Journal.

Vol. 7
1979

Annual Subscription Rates (1980)

Published bi-monthly. For libraries, university departments, government laboratories, industrial and other multiple-reader institutions U.S. $143.00 per annum; 2–yr subscription rate $271.70 (including postage and insurance). *Specially reduced rates to individuals*: In the interests of maximising the dissemination of material published in this important international journal we have established a two-tier price structure. Any individual whose institution takes out a library subscription may purchase a second or additional subscription for their use at the much reduced rate of U.S. $30.00 (£12.50) *per annum.* Subscription enquiries from customers in North America should be sent to: Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, NY 10523, U.S.A. and for the remainder of the world to: Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England.

Microform Subscriptions and Back Issues

Back issues of all previously published volumes are available in the regular editions and on microfilm and microfiche. Current subscriptions are available on microfiche simultaneously with the paper edition and on microfilm on completion of the annual index at the end of the subscription year.

Publishing Office

Pergamon Press Ltd., Hennock Road, Marsh Barton, Exeter, Devon EX2 8RP, England

Subscription and Advertising Offices

Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England
Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, USA

Copyright 1979 Pergamon Press Ltd

It is a condition of publication that manuscripts submitted to this journal have not been published and will not be simultaneously submitted or published elsewhere. By submitting a manuscript, the authors agree that the copyright for their article is transferred to the publisher if and when the article is accepted for publication. However, assignment of copyright is not required from authors who work for organisations which do not permit such assignment. The copyright covers the exclusive rights to reproduce and distribute the article, including reprints, photographic reproductions, microform or any other reproductions of similar nature and translations. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, electrostatic, magnetic tape, mechanical, photocopying, recording or otherwise, without permission in writing from the copyright holder.

U.S. Copyright Law applicable to users in the USA

The Article Fee Code on the first page of an article in this journal indicates the copyright owner's consent that in the USA copies may be made for personal or internal use provided the stated fee for copying beyond that permitted by Section 107 or 108 of the United States Copyright Law is paid. The appropriate remittance should be forwarded with a copy of the first page of the article to the Copyright Clearance Centre Inc. PO Box 765, Schenectady, NY 12301. If a code does not appear copies of the article may be made without charge, provided permission is obtained from the publisher. The copyright owner's consent does not extend to copying for general distribution, for promotion, for creating new works or for resale. Specific written permission must be obtained from the publisher for such copying. In case of doubt please contact your nearest Pergamon office.

Printed in Great Britain by A. Wheaton & Co. Ltd, Exeter

omega

The International Journal of Management Science
Volume 7, Number 1, 1979

Contents

Samuel Eilon	1	Editorial—A Question of Size
WH Jones, MH Oakley	9	A Case Study in Venture Management
JRC Wensley	15	Relative Success of Marketing Decision Aids: A Survey
R Flavell, E Penn, GR Salkin	25	The Planning of a Department Store
BHP Rivett	33	Futures Literature and Futures Forecasting—A Critical Review
Dudley Jackson, Tim Sweet	43	The World Strike Wave
Daniel J Brown	55	Modelling the Use of the Manager's Time
RJ Graham, J Seltzer	61	An Application of Catastrophe Theory to Management Science Process
Lawrence A Gordon, David F Larcker, Francis D Tuggle	67	Informational Impediments to the Use of Capital Budgeting Models
	75	**Memoranda**
Stephen P Bradley, Sherwood C Frey Jr	77	Capital Rationing and Discounting in Public Sector Investment Decisions
John F Kottas, Hon-Shiang Lau	80	A Decision Model with Stochastic Cost and Demand Curves
George Cosmetatos, Nicholas Barbaroussis and Samuel Eilon	82	The State of Variable Costs in Profitability Planning
	85	International Register of Theses and Dissertations
	i	Announcements

Special Notice

To avoid simultaneous, multiple submissions to journals having similar interests, the Editors of *Management Science*, the *Operational Research Quarterly*, *Operations Research*, *Omega* and *Naval Research Logistics Quarterly* have developed a system of exchange submission lists.

Because refereeing capability is a scarce resource, we cannot tolerate the duplication of effort that follows from multiple submissions. We hope that the elimination of multiple submissions will make the refereeing process more efficient, and thereby decrease the time required for refereeing. This is a major objective of all five publications. Other journals with similar interests may soon be included in this information interchange.

Vol. 7
1979

omega

The International Journal of Management Science
Volume 7, Number 2, 1979

Contents

Editorial

SAMUEL EILON 89 Ackoff's Fables

Feedback

SIR TERENCE BECKETT 101 People and Size

JUDITH ATHERTON, GORDON WILLS 104 Information as a Tangible Benefit

Main Papers

RUSSELL MOSELEY 105 Technical Change and Employment in the Post-War Gas Industry

RICHARD M BURTON, WILLIAM W. DAMON 113 Budgets for Integrating Decentralized Investment and Production Planning

COLIN EDEN, DAVID SIMS 119 On the Nature of Problems in Consulting Practice

MICHAEL FIRTH 129 The Disclosure of Information by Companies

SUBHASH NARULA, UGONNAYA OGBU 137 An Hierarchal Location–Allocation Problem

ROBERT W HAESSLER 145 Solving the Two-Stage Cutting Stock Problem

PREM VRAT, SUBASH BABU 153 Inventory Control of Spares in a Transport Corporation—A Case Study

Memoranda

HP WILLIAMS 163 Four Examples of Artless Modelling

SAMUEL EILON 166 Some Useful Ratios in Evaluating Performance

International Register of Theses and Dissertations 169

omega

The International Journal of Management Science
Volume 7, Number 3, 1979

Contents

Editorial

SAMUEL EILON 173 Lateral Consistency?

Feedback

RL ACKOFF 179 Ackoff's Fables

STEN OLOF GUSTAVSSON 179 A Question of Size

NANCY FOY 180 Political Man

CA GLISSON 180 In Defense of Determinism

Main Papers

PETER G MOORE 183 Handling Financial Risks

RALPH L KEENEY, RAM B KULKARNI KESHAVAN NAIR 191 A Risk Analysis of an LNG Terminal

JEAN-CLAUDE LARRÉCHÉ, EDWARD C STRONG 207 A Supportive Application of Management Science to Marketing Strategy Formulation

RA BRYER 219 The Status of the Systems Approach

JR KING 233 Scheduling and the Problem of Computational Complexity

JAMES W GENTRY, JAMES SHANTEAU 241 Revision of Consumer Intention: Two Models

STANFORD BAUM, ROBERT CARLSON 249 On Solutions that are Better Than Most

LEWIS J CORNER 257 Linear Programming: Some Unsuccessful Applications

International Register of Theses and Dissertations 263

O.M.E. 7/3—A

Vol. 7
1979

omega

The International Journal of Management Science
Volume 7, Number 4, 1979

Contents

Editorial

SAMUEL EILON 269 Power Corrupts

Feedback

ROY AMARA 273 Planning, Futures, and the Skeptics

Main Papers

DUNCAN DAVIES 275 Change, Clarity and Clausewitz

DALEN T CHIANG 287 Liquidity Planning in a Small Bank

MARTIN WALBANK 297 An Integrating Model of Motivated Work Behaviour

MO GIESEN, PG BENNETT 309 Aristotle's Fallacy: A Hypergame in the Oil Shipping Business....

STEVEN NAHMIAS, GEOFFREY LEVINE 321 Inventory Management and Cost Benefit Analysis: A Case Study

NIGEL MEADE, MALCOLM J FINNEY 333 Analysis of the Choice Between Domestic and Foreign Loan Facilities

LUC A SOENEN 339 A Portfolio Model for Foreign Exchange Exposure Management

ARTHUR J KEOWN, BERNARD W TAYLOR III, CALVIN P DUNCAN 345 Allocation of Research and Development Funds: A Zero-One Goal Programming Approach

Memoranda

SH ARCADE 355 An Exercise in Probability Encoding

N DOHOLAKIA, R KHURANA, L BHANDARI, MN VORA 359 On Normative Policy Models in Developing Countries

WILLIS R GREER JR, GERALD J LACAVA 361 Normal Approximations for the Greater of Two Normal Random Variables

RALPH L KEENEY 364 Public Risk From Potential LNG Spills

NK KWAK, MJ SCHNIEDERJANS 367 A Goal Programming Model for Improved Transportation Problem Solutions

DRF SIMPSON 371 The Efficient Use of Non-Renewable Resources

International Register of Theses and Dissertations 375

OME 7/4 A

omega

The International Journal of Management Science
Volume 7, Number 5, 1979
Special Issue on Public Sector Decision-Making

Contents

Editorial

DAVID PEARCE	379	Public Sector Decision-Making

Main Papers

MS COMMON	385	External Costs: Myth or Rationale for State Intervention?
KJ BUTTON	399	Models for Decision-Making in the Public Sector
RK TURNER	411	Cost–Benefit Analysis—a Critique
J LOWE, D LEWIS	421	Economic Approaches to Environmental Decision-Making
M BARRETT	431	Environmental Impact Statements
CA NASH	441	The Use of Economic Appraisal in Transportation Analysis
ALASTAIR GRAY, ROBERT STEELE	451	Programme Budgeting in the Health Sector
JOHN M TREDDENICK	459	Cost-Effectiveness in Defence Expenditure
DAVID HEALD	469	Public Expenditure and Public Accountability

Vol. 7
1979

omega

The International Journal of Management Science
Volume 7, Number 6, 1979

Contents

Editorial

SAMUEL EILON	481	Evaluating Management Education

Feedback

KC BOWEN	489	A Question of Size
JOHN J MORSE	489	The Vroom–Yetton Leadership Model
MARTIN K STARR	490	Ackoff's Fables

Main Papers

ANDREW STRATTON	493	Energy Forecasting
KC BOWEN	503	Personal and Organisational Value Systems: How Should we Treat these in OR Studies?
JR MANN, GH MITCHELL	513	Industrial Relations Problems—OR's Contribution
MYLES A VOGEL	521	Increasing the Effectiveness of Modeling
MICHAEL J GINZBERG	527	Improving MIS Project Selection
JPC KLEIJNEN	539	Evaluation of Management Information Systems
WILBUR A STEGER	545	Assessment of Fifteen Years of Urban Modeling
PANDU R TADIKAMALLA	553	The Lognormal Approximation to the Lead Time Demand in Inventory Control
DOMINIC SCULLI	557	Dynamic Aspects of Line Balancing

Memoranda

STEPHEN E BECHTOLD	565	Quantitative Models for Optimal Rest Period Scheduling: a Note
HIROHIDE HINOMOTO	567	Capital Rationing Decision with Internal and External Funds
LARRY E SHIRLAND, PETER M ELLIS	569	Optimum Location of Training Centers
RN WOOLLEY	574	Creativity and Operational Research
International Register of Theses and Dissertations	577	
Errata	i	

OML 7 6- A

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 1–7
0306-0483/79/0201-0001$02.00/0
© Pergamon Press Ltd. 1979. Printed in Great Britain

EDITORIAL
A Question of Size

EVER SINCE Schumacher published his memorable book 'Small is Beautiful' [5], industrialists and economists have been arguing the merits of his case that large, monolithic and highly centralized organizations are not only cumbersome to manage, but have become undesirable, both on economic and social grounds. The prosperity of large industrial companies in the US has always been regarded as tangible evidence of the economies of scale, through the ability to spread central management overheads and high fixed costs of sophisticated production facilities over a large volume of output, and thereby pose a real threat to competitors operating on a smaller scale.

The nationalization of certain key industries in Europe and the creation of the Common Market, provided opportunities for very large organizations to be set up on this side of the Atlantic, and spasmodic fevers of mergers and acquisitions during the past two decades have reflected the assiduous conviction that operations on a large scale provide greater opportunities for systematic planning and for greater efficiency.

As Schumacher comments, "the great majority of economists and business efficiency experts supports this trend towards vastness", but he adds: "In contrast, most of the sociologists and psychologists insistently warn us of its inherent dangers—dangers to the integrity of the individual where he feels as nothing more than a small cog in a vast machine and where the human relationships of his daily working life become increasingly dehumanised; dangers also to efficiency and productivity, stemming from ever-growing Parkinsonian bureaucracies" [5, p. 225]. He is further concerned about the impact of large companies on the environment, and he argues: "Small-scale operations, no matter how numerous, are always less likely to be harmful to the natural environment than large-scale ones, simply because their individual force is small in relation to the recuperative forces of nature. There is wisdom in smallness, if only on account of the smallness and patchiness of human knowledge, which relies on experiment far more than on understanding" [5, p. 31].

Thus, worries about size stem from several issues relating to the long term impact of the organization, on its own performance and on its environment. First, there are people who question the fundamental proposition that economies of scale always lead to improved performance and efficiency. Admittedly, certain fixed costs per unit of output tend to decline as the volume of output increases; similarly,

central purchasing, unified distribution networks, integrated and coordinated processing, corporate planning and financial control—they all appear to have a greater significant effect on the organization as the scale of operations is increased, and they all tend to contribute to a reduction in unit costs. But size and complexity impose strain on the information system, many people need to be informed or consulted about the implementation of new decisions or about changing procedures, so that the cost per unit of output of 'keeping the organization going' may well increase with size. It is therefore often argued that the resultant total unit cost function, when plotted against volume of output, assumes a U-shape, giving rise to the notion that every industrial activity has its optimum size with respect to total unit cost. Hard empirical evidence on this score is somewhat difficult to come by, largely because of problems in measurement, in interpreting financial data, and in comparing like with like, so that several investigators (for example, Gold in his numerous studies of long term cost curves [4]) have become rather critical of the conventional wisdom propagated in text books regarding the theory of the firm. What is perhaps becoming clear is that evidence of economies of scale in one industry need not give any guide as to the effect of scale in another, and that not only technology, but the market and the environment may prove to be major determinants.

Vol. 7
1979

Secondly, even if the advantage of scale in pure cost terms is not in doubt, the effect on competition, and hence the long term effect on the market, can have serious consequences. As an enterprise becomes larger and more powerful, it tends to inhibit its competitors, it discourages newcomers into the field, and it increases the probability of further entrenchment by fewer companies in the market place. Large enterprises have been known to take over small competitors, or to drive them out of business, and if it had not been for anti-trust legislation and governmental watchdogs we would have probably witnessed a more rapid growth of monopolies, increasingly dominating the supply of goods and services, the control of prices and the range of choice available to the customer. In such circumstances, the initial economic advantage of scale soon evaporates, in that large organizations become sufficiently strong to be impervious to criticism and immune from attack in the market; as a result, they begin to harbour bureaucracy and inefficiency, with detrimental consequences to the customer.

Thirdly, I have already alluded to the effect of scale on the workforce. The 'cog-in-the-wheel' feeling is enhanced when the employee suspects that his contribution to the well-being of a vast organization is rather minimal, that his identity is lost in the crowd, that he has little room to exercise his skill and judgement, or that his grievances are not likely to reach the senior echelons of the enterprise. Size tends to impede communications, to enforce a sense of isolation and to foster frustration. It is interesting to note that some large companies

like to operate under a single trade-mark, and promote the notion of a unity of purpose by encouraging all employees (even when they are widely dispersed geographically) to identify themselves with the corporate image (typical examples are oil companies, computer manufacturers, and airlines), whereas other companies trade under several names and believe in the advantages of employees clinging to localized (or product-related) identities, each group developing its own *esprit de corps* (in some cases, the general public, and even many of the employees, are not aware of the fact that companies trading under different names belong to the same group). Admittedly, the product and market considerations may be decisive factors in determining whether an enterprise should have a single or several identities, but there are many cases where either mode of operation could be viable, and the management decision to opt for one rather than the other is to some extent a reflection of their view of the effect of scale on human and industrial relations.

Fourthly, as Schumacher argues, size is important in considering the effect on the environment. If a small enterprise fails or misbehaves, the consequences can be more easily absorbed or tolerated by the environment than when a large enterprise is involved. This simple fact has been repeatedly demonstrated in recent years by various disasters, such as the explosion of large chemical plants, the break-up of large oil tankers, the failure of large electricity supply networks, and so on. But apart from the physical effects of fire, pollution and contamination that such disasters cause, there are the far-reaching economic consequences to a locality, or even to a whole country, when a large enterprise experiences a financial or organizational collapse, and this is a worry that is particularly highlighted in the case of multi-nationals, where a change of fortune and the impact of decisions made at some remote head office add to the distress of a helpless national government in its attempt to control the economic scene in its own country [3].

These four aspects of the effect of size touch upon some of the problems that are involved. Economic performance and efficiency of operations are clearly of prime importance, but they cannot be divorced from other considerations, not only because all the other factors that need to be examined are ultimately bound to have economic consequences, but because those aspects of the problem that cannot be expressed in purely financial terms still have a validity of their own, which managers and society cannot ignore. Furthermore, it is evident from the foregoing discussion that there are many interests involved—interests of top management, employees, labour unions, customers, suppliers, competitors, financial institutions, national governments; and since many of these interests are in conflict, it is necessary to identify the point of view from which the analysis is embarked upon before any conclusions can be drawn.

A typical range of problems in this field are cited by Cantley and Glagolev [2]:

> (i) "a traditional economic model, embodying relationships between size, performance and cost, with the resulting historic trend towards increasing size for maximum economic advantage;
>
> (ii) a concern with flexibility or robustness in the context of environmental change;
>
> (iii) management, estimation (of cost and time), and control problems in the creation or installation of very large organizations or units of plant;
>
> (iv) new problems of management and control arising in the operation of large organizations and units of plant;
>
> (v) increased problems of statistical reliability or security which are accentuated by the concentration of capacity into fewer, larger centers."

This is by no means the whole story. Apart from the fact that it is confined to the managerial viewpoint and that similar lists can be compiled to reflect the concern of other interested parties, there are many further aspects of economic performance and managerial control that need to be considered, and some of the issues are rather complex and clearly indicate that the question of scale can hardly be settled by an over-riding single criterion.

Vol. 7
1979

Take, for example, the implication of size on motivation and the performance of the individual. Schumacher highlights the inevitable conflict in organizational design between *order* and *freedom*, and yet the need for an enterprise to have both. On the one hand the organization has to maintain orderliness, consistency in behaviour, clarity of procedures; such a state of corporate discipline is achieved through the development of programmed decisions and the compilation of a rule book, which is nourished and sustained by a creeping bureaucracy. At the same time, the enterprise needs to encourage initiative and to provide for new developments through innovations and radical reappraisals. As Schumacher puts it: "The larger the organization, the more obvious and inescapable is the need for order. But if this need is looked after with such efficiency and perfection that no scope remains for man to exercise his creative intuition, for *entrepreneurial* disorder, the organization becomes moribund and a desert of frustration" [5, p. 227].

He therefore postulates five principles for the management of large organizations, based on preserving the identity of small sub-units [5, pp. 228–234]:

> "1. *The principle of the subsidiary function*, to protect the lower levels of the organizational hierarchy from the stifling effects of over-centralization, in the belief that if sub-units maintain freedom to act and are held responsible for their actions then "the centre will gain in authority and effectiveness".

2. *The principle of vindication*, to judge the performance of the subsidiary function against given criteria, and to defend it from unnecessary reproach.

3. *The principle of identification*, to emphasize the unique identity of the sub-unit, for example by insisting on its having its own profit and loss account, even when its use is limited to internal purposes.

4. *The principle of motivation*, to ensure that individuals get a sense of satisfaction from their contribution to the organization, and to combat "the relative incomprehensibility that stems from its very size".

5. *The principle of the middle axiom*, to strike a balance between order and entrepreneurial activity."

It is against this background that investigators in this field are bound to be intrigued by what captains of industry have to say about the effect of size on the operations of their companies. After all, they are the people who have the experience of running such organizations and struggling with problems that constantly emerge; they should therefore be in a unique position to provide illuminating accounts of how they set about evaluating the benefits and disbenefits of scale.

Such accounts are, alas, few and far between. Not many senior executives feel compelled to commit to print the results of detailed analyses of large scale operations that they may have carried out. It was, therefore, with particular interest that I turned to a paper [1] written by TN Beckett (now Sir Terence), the Chairman and managing Director of the Ford Motor Co. in the UK, a company with 70,000 employees. The title of the paper, 'Problems of size and efficiency in industry today', seems precisely to describe the subject matter that we are concerned with here, and I looked forward with eager anticipation to see what further light the paper could shed on this discussion.

The author starts promisingly enough: "If... we conclude that some large scale companies are necessary, then I want to go on to ask what do we have to do to enable them to be successful, with all the problems we have in many large companies with industrial relations". He then emphasizes the importance of technology: "The differences in production cost between one technology and another are so large, in some cases, that no amount of diligence or changes of scale using a given technology can make the costs competitive with those available from a superior technology", adding that in one particular case a cost comparison between his company and a competitor producing one third of the volume of a similar model, the cost difference was so significantly in favour of the large producer, that it could

threaten the very existence of the more costly operation. "Scale, therefore, is essential to survive and succeed in some industries", and "it is certainly the case in the motor industry", where Beckett believes that the optimum size for an assembly plant in Western Europe today is for about 300,000 cars a year and about two million for engines.

The author admits that "other things being equal, the smaller the company, the easier it is to manage, and within a large organization one should try to break it down to smaller more manageable, more human units", but he gives no indication as to how this can be done in a large firm in the motor industry. If, as he maintains, the manufacturing activities in that industry are dominated by the prevailing technology, and if many of the operations are sequential, then it may not be possible to employ the managerial principles postulated by Schumacher. In any event, Beckett gives us no clue as to their relevance.

As for the human problem in a large firm such as his, he says: "The received wisdom begins as follows. Mass production means boredom, monotony, unpleasant machine-paced work and alienation. What can we say about these? We have in my Company, in the last twenty-five years, since I worked on the shop-floor, made many improvements in working conditions and in reducing monotony. But the fact is that unless we want to reduce our material aspirations, both individually and socially in this country—and I don't see any evidence at all that we do—the present state of our technology is such that the only way our wealth can be produced at the present time is with some boring jobs. We shall improve them as time goes on, but some jobs *are* boring at the present time". (Preachers on self-actualization, please note!)

Vol. 7
1979

The dilemma, then, remains. If size is essential for survival at least in some industries, all other considerations are of secondary importance. Communication, alienation, industrial relations, motivation, environmental effects—they are all irritants that need to be attended to as problems arise, and "we shall improve them as time goes on", but they must not be allowed to distract the organization from its prime objective. The acceptance of size as an inviolable constraint therefore implies that no satisfactory and lasting cures can be found for some of the fundamental complaints alluded to earlier, except perhaps for the approach advocated by Schumacher, which may not be universally applicable. And yet, size harbours its own threat to survival. As size increases (and what is optimum in economic terms today may well be regarded as too small tomorrow), and as operations within the organization become more interdependent and sequential, there is the increasing risk that a failure of one sub-system, however small, will bring the whole enterprise to a halt. A chain is as strong as its weakest link, and in order to reduce the impact of its failure, there is a need in industry for alternative links, and above all for

some slack. Elimination of slack, which is one of the prime aims of large organizations, is obviously beneficial from an economic viewpoint in the short term, but this benefit needs to be weighed against a massive breakdown and its potential horrendous consequences for all the people who are likely to be affected. It is time that captains of industry came up with some convincing answers as to how their operations, and society as a whole, can be properly protected.

SAMUEL EILON
Chief Editor

REFERENCES

1. BECKETT TN (1977) Problems of size and efficiency in industry today. *The Roy. Soc. of Arts Jl*, **CXXXV**(5254), 624–636.
2. CANTLEY NF & GLAGOLEV VN (1978) Problems of scale—the case for IIASA research. **RM-78-47** (Sept. 1978), International Institute for Applied Systems Analysis, Austria.
3. EILON S (1973) Serving people and nations everywhere. *Omega* **1**(5), 531–535.
4. GOLD B (1971) *Explorations in Managerial Economics*. Macmillan, London.
5. SCHUMACHER EF (1973) *Small is Beautiful*. Blond and Briggs, London.

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 9–13
© Pergamon Press Ltd. 1979. Printed in Great Britain
0306-0483/79/0201-0009$02.00/0

A Case Study in Venture Management

WH JONES
MH OAKLEY

University of Aston Management Centre, Birmingham, UK

(*Received July* 1978)

This paper examines the venture concept as a means of stimulating new product development within large companies. The literature reveals a significant difference between the venture idea as originally conceived in North America and as applied in European companies. A case study of a UK venture experiment is then presented which clearly demonstrates the resistance to substantive change which may characterise large organisations.

INTRODUCTION

INDUSTRIAL venture systems have been devised and implemented for many years, but examination of the literature dealing with the concept shows a lack of consistency of definition. The 'venture' label is freely applied to management styles which differ in important respects. Often, the only common feature is that the object of discussion is an industrial subgroup concerned with innovation. In particular, there is a marked failure to distinguish between venture systems in the USA and the quasi-venture systems which tend to operate elsewhere.

DEFINITION AND FEATURES OF THE VENTURE CONCEPT

The venture concept is based on the theory that big companies stand a better chance of successfully launching new business enterprises if they can create conditions, devoid of bureaucracy, where the motivations of an entrepreneur can be given full scope. In essence, venture systems attempt to simulate the incentives and challenges that small businesses present to those who run them, particularly the prospect of rapid self-advancement. Thus, the most general definition of what is meant by the venture concept of organisation is "the formation of new businesses by tapping existing technical and human resources. It is an innovative form of management that is designed to get away as much as possible from red tape inherent in a large organisation" [8].

A venture team is usually set up under the control of a manager who is not restricted in his approach by the organisation of the parent company. Typically the manager and team members are young people who have already demonstrated above-average ability. This is an essential feature of venture groups. Simulation of a small business environment is reinforced in the US where it is common for all members of the team to invest substantial amounts of their own funds in addition to the company contribution [1]. It is likely to be the first opportunity team members have to take such a business risk and their endeavours for success and reward can be expected to be of a high level. On this side of the Atlantic, in line with a more conservative business approach, investment by team members is not common. The main incentive is the prospect of early advancement and the opportunity to run a business at an early stage. This difference in approach is very important and will be referred to again.

Although it is claimed that venture techniques are widely used in many large firms (such as Monsanto, Du Pont and General Electric in the States [4] and British Oxygen [3] and ICI [7] in the UK) there is little evidence in the literature of attempts to compare

systems and quantify results. Invariably, papers dwell on business successes and make only passing reference to the features of the organisation responsible for achieving the success.

However, a useful paper has been published by Hill and Hlavacek [5] which reports on a study of over 100 venture teams in large companies in the United States. By using questionnaires and personal interviews, a number of unique characteristics were described which distinguish venture teams from traditional forms of industrial organisations. A list of these characteristics helps to define more specifically the nature of venture organisations:

(1) *organisationally separate*—Burns and Stalker [2] found that the most successful new product ventures were geographically as well as organisationally separate;

(2) *multidisciplinary*—the interaction of team members with different skills was considered essential for creativity;

(3) *diffusion of authority*—team members were not confined by job descriptions and were free to develop working relationships; team managers were often indistinguishable from other members;

(4) *environment of entrepreneurship*—"free wheeling" atmosphere together with financial or personal commitment by team members;

(5) *top management support*—venture team members usually report to managers at the highest level;

(6) *market oriented missions*—rather than product oriented, but otherwise broad discretion in method of operation;

(7) *flexible life span*—lack of strict deadlines for the accomplishment of tasks or goals.

In his report on a business venture in ICI, Vernon [7] covers all these points. In some cases there is qualification or emphasis which one might expect to distinguish a British study from an American one. For example, he tempers the concept of a "free wheeling" atmosphere with comments about the need for plans to be made carefully and all staff to be "finance-conscious".

If team members have a personal financial commitment, this problem will be self regulating. But where it is common, as in British companies, for all finance to be channelled from the parent company, it is necessary to have checks and procedures to ensure spending is kept to reasonable limits. Unfortunately, this intrusion of bureaucracy, however slight, can seriously endanger the venture philosophy. Perhaps this is why Vernon stresses the need for the venture manager to create and maintain optimism within the team. Despite this, however, Vernon presents a generally hierarchical view of venture organisation. In contrast, Hill and Hlavacek stress the need for the leader to be one amongst equals.

A RECENT QUASI-VENTURE EXPERIMENT

The experiment took place in a subsidiary of a large British group and we shall refer to this subsidiary as the XYZ Company. For many years XYZ had produced a range of small to medium volume products which were supplied to industrial users. Substantial innovation had occurred after World War II and a number of important and successful products were launched.

Vol. 7
1979

Most of these were derivatives of existing products and they satisfactorily served to enlarge existing markets and also to create new ones. One major new product was quite radical and established the company in a useful new business. The product was developed outside the company in a project funded jointly by XYZ and the 'Customer'. It is reported that the success of this project resulted from the determination of three forceful personalities—the inventor of the product, a senior manager of XYZ and a senior manager of the 'Customer'. The product was highly profitable for a few years but then turnover began to decline.

The decline of this product formed part of a general decline in XYZ's turnover in the period from the mid-1960s to the present. In the ten years up to 1976, turnover had decreased in real terms by 25 per cent. With improvements in production processes, employment had reduced by some 50 per cent in the same period, causing the loss of over 2,000 jobs. Many of the best markets had disappeared and some products had become obsolescent. Those which remained were mainly being sold to stable or declining markets. To a large extent the company's fortunes were dependent upon the success or failure of a handful of major customers to whom much of the production was sold.

During the years of decline, XYZ had made some efforts to introduce new products but none had met with commercial success. A reason for this was that the management team had come to be dominated by production managers who saw the problem in terms of cost cutting and improvement to existing manufacturing processes. The innovative flair which had flourished during the 1950s had been submerged as the products which it created moved into the mature stages of their lives. Many creative individuals had left the company to work elsewhere, or had retired, or had resigned themselves to grow old with their products.

There is evidence that during the period from 1966 to 1976, expenditure on product innovation was grudging and inadequate. To quote the percentages of the company turnover which were spent on research and development and to compare these with the published average figures for the industry, would enable an astute reader to make guesses at the real identity of XYZ. To support the argument it must merely be stated that research and development expenditure in XYZ ran at about one third the average level maintained by similar British firms in the industry.

In addition to this financial perspective, interviews with members of staff and inspections of written records confirmed the trend. It was found that the innovative projects which were allowed to start up, were under constant attack by critics. They were treated as irrelevant even by the most senior members of the company.

In 1976 some new senior managers began to make their mark in the company. Amongst a number of changes which took place, was a decision to devote more resources to a programme of new product development. It was decided that the programme would be run on venture lines and a new division of the company was formed to provide a framework for the development projects. This new division was created by the reorganisation of the existing production divisions of the company. See Figs 1 and 2 for simplified organisation diagrams.

It was intended that each venture project would be run as an independent enterprise. Its manager would report straight to the new division's director, who, together with the other divisional directors, was a board member. In this way, it was anticipated that decisions about appraisal and finance would be taken speedily and that this high level linkage would ensure general "protection from the system".

The grouping of the new ventures with the old division A was an attempt to use the experience built up in that division. The ventures were financially separate with independent budgets and would buy services and goods from division A in the same way as from external sources. As the ventures flourished it was thought that each would become established as separate operating divisions in their own right.

Six months after the structural changes, it was clear that the new ventures were not working as expected. The first of the projects had been run down and most of the staff posted elsewhere. Other projects were failing to prosper and their managers and staff were becoming demoralised. The following six factors were seen to be responsible.

(1) *Inadequate initial screening of projects and lack of marketing awareness*

Such was the initial enthusiasm of the management of XYZ for the venture idea that the first product for development was selected without proper consideration of its potential. Stevens [6] claims that most venture failures arise because neither the product nor the manager is sufficiently outstanding. In too many cases, he says, the product does not have either a sufficiently large sales potential or a big enough profit margin to provide the foundation for an independent business. On the basis of the information available, XYZ could not have forecast sales or profit with any certainty.

Perhaps even more fundamental was the failure to be sufficiently market-oriented in the selection of projects. XYZ was particularly weak in terms of marketing and much previous innovation had failed because technical ingenuity had taken precedence over gaining an understanding of the user's needs. Several

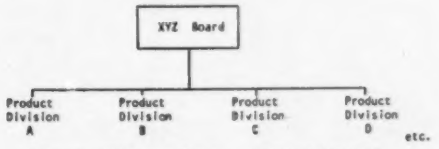


FIG. 1. *Old organization of XYZ Limited.*

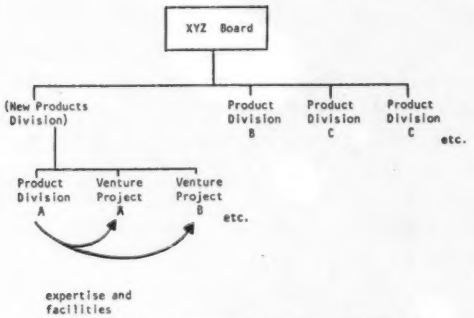


FIG. 2. *New organization of XYZ Limited.*

excellent products had been devised for which it was eventually found that only insignificant sales could be obtained. This does not mean that no results of speculative development work should be considered. Vernon indicates that the success of the ICI venture operation stemmed from a chance technical discovery; but the starting point was a market analysis and only when an opportunity was identified did a search begin for suitable technology. These points are also reinforced by Gardner [3].

(2) *Lack of commitment by top management*

Faced with the news of technical problems with the first project, the reaction of the management was to take resources away from it. Conversely, when progress and prospects looked good, ever more optimistic targets were set. The point here is that once a potential project has been thoroughly analysed and then selected, the venture manager should be insulated from inconsistency. Whilst the company was naturally anxious to see projects succeed, it should have been ready to accept that some—even a majority—might fail, and the best chance of success would be assured by steady support.

(3) *Lack of protection from the organisation*

Very soon after the first venture project was established, the pressure of bureaucracy began to be felt. At the corporate level, the Strategic Planning Department decided it would need the venture manager to provide extensive product appraisal information. A lengthy questionnaire was issued which demanded information in great detail about technical and commercial aspects of the project. It was returned for revision several times and the estimates which it contained began to be regarded as facts. More disturbing, the simple predictions, acknowledged to be 'guestimates' upon which the project had been established were increased without good reason during the process, so that the project appeared to grow more attractive and expectations were raised.

At the Divisional level, the Buying Department monitored the purchases of material and equipment made by the venture team members. It complained that through inexperience, high prices were being paid. It insisted on the use of 'correct procedure' and the result was much paperwork and long delays instead of immediate deliveries.

The effect of these actions was to divert and dilute the entrepreneurial energy of the team members. A better 'selling job' of the venture concept to the rest of the company by the management may have helped, but short of geographical separation, it is unlikely that established service departments could have stood aside, perhaps to be eclipsed if the new venture grew to success.

(4) *Reluctance to give venture manager freedom of action*

The problems concerning buying have already been mentioned. There was also a tendency for staff to be appointed to the projects who were not the choice of the venture manager. In the area of finance, although budgets were agreed and allocated, the venture manager was required to obtain authorisation for even relatively small expenditures. The effect of these interferences and restrictions was to blunt the enthusiasm of the manager and his team.

(5) *Tendency for mechanistic forms of operation to be created within the venture groups*

The chosen venture managers had gained their experience largely within the parent company, and it is not surprising, therefore, that they tended to develop their groups along traditional lines. Hanan [4], Gardner [3] and Hill and Hlavacek [5] all suggest that an informal organisation with uninhibited communications is the ideal form for innovative groups. The groups set up by XYZ soon showed the manifestations of an emerging hierarchy, such as the grading of offices, competition for appointment of secretaries and a general growth of paperwork. Inevitably, preoccupation with these ele-

Vol. 7
1979

ments led to reduced vitality in the real work of the project.

(6) *Lack of real incentive for team members to succeed*

In this venture experiment there was no question of financial investment by the team members, to be followed by the prospect of high reward when the project became a success. The main incentive lay in the suggestion that if all went well and a good business was created, then the participants would assume positions at a much higher level than they could otherwise have expected. However, another possibility which was stated was that as the project developed, it would be 'grafted' on to an existing operation and come under the control of an experienced senior line manager.

CONCLUSIONS

Appraisal of the potential of the venture concept is made difficult because the published research tends to be concerned with projects that have been commercially successful. Hence, it is felt that this paper will help to increase understanding of the venture concept by presenting a case study of an unsuccessful attempt to introduce this concept within a large UK company and by diagnosing organisational factors that may render a venture group impotent.

The unsuccessful XYZ experiment highlights a number of essential features which are necessary if a venture or quasi-venture exercise is to achieve its objectives. The essential features which were seen to be absent from XYZ can be summarised as:

(1) proper understanding of products and markets;

(2) consistent top management support;

(3) protection from large company bureaucracy;

(4) venture manager's freedom of action;

(5) free wheeling, informal operation;

(6) real incentives for team members.

The literature review indicated the importance of the difference between the operation of the venture concept in the US and elsewhere. The original concept was tied up fundamentally with financial participation by team members as well as by the parent organization and third parties. This element of personal commitment is not usual outside the US and was not evident in this case study; consequently the results obtained by quasi-venture groups can be quite different from those obtained by real venture groups.

It is dangerous to generalise from a single case but the XYZ exercise suggests that two important aspects require further study:

(i) *the need for geographical separation*—many of the problems encountered by XYZ were aggravated because the groups had to operate within an existing company location. This made it too easy for the established bureaucracy to interfere and for conspicuous entrepreneurial action to be criticised.

(ii) *the need for real incentives for team members*—whether this is by bonus payments, equity participation or some other means, the requirement is that there should be the prospect of relatively considerable personal gain. Indication of rapid promotion is not, in itself, sufficient incentive.

REFERENCES

1. Ben Daniel DJ (1974) The technical ventures operation, in *Technology Transfer: Proceedings of the NATO Advanced Study Institute on Technology Transfer* (Eds Davidson HF, Cetron MJ & Goldhar JD). Noordhoff Publishing, Leiden, The Netherlands.
2. Burns T & Stalker GM (1966) *The Management of Innovation* (2nd Ed.), Tavistock Publications, London.
3. Gardner JB (1973) Innovation through new ventures: new venture concept in BOC. *R & D Mgmt* **3**(2), 87–89.
4. Hannan M (1969) Corporate growth through venture management. *Harv. Bus. Rev.* **47**, 43–61.
5. Hill RM & Hlavacek JD (1972) The venture team: a new concept in marketing organisation. *Jl Mktg* **36**, 44–50.
6. Stevens HAR (1972) Criteria for selection of venture capital projects. *R & D Mgmt* **3**(1), 41–45.
7. Vernon A (1974) A new business venture. *R & D Mgmt* **4**(2), 85–88.
8. Weintraub PI (1975) Big business goes small. *Bus. Q.* **40**(3), 48–55.

ADDRESS FOR CORRESPONDENCE: *Dr W H Jones, The University of Aston Management Centre, 158, Corporation Street, Birmingham B4 6TE, UK.*

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 15–23
© Pergamon Press Ltd. 1979. Printed in Great Britain
0306-0483/79/0201-0015$02.00/0

Relative Success of Marketing Decision Aids: A Survey

JRC WENSLEY

London Graduate School of Business Studies

(*Received July* 1978; *in revised form August* 1978)

A number of analytical techniques have potential applications in marketing management. A questionnaire survey of marketing managers reveals that although the level of awareness of such techniques is high the usage made of many of them is low. Optimising techniques are generally less widely used than those concerned with aggregation of judgements. Associations in the usage of techniques appear and although limited differences between types of industry are apparent in this respect, organizational differences appear to be less significant.

THIS paper is concerned with the results of a detailed questionnaire sent to 60 middle and senior personnel with a very substantial bias towards the line marketing function: 41 provided usable replies. This questionnaire as well as a number of detailed interviews formed a programme of research discussed previously [5].

OBJECTIVES AND USAGE

The major objective of the research was to ascertain what decision aids were actually extensively used by line managers and to evaluate possible implications of these results. This article will where appropriate contrast the impressions gained as a result of the detailed interviews with the survey results.

To ensure consistent definitions for the various techniques, a short description of each technique was circulated with the initial questionnaire. The definitions used are given in Table 1. Tables 2 and 3 indicate the overall results for the whole sample. Table 2 gives the actual response figure ($n = 41$) whereas Table 3 gives percentage figures calculated on different appropriate bases as indicated. Table 4 is a rank ordering by usage. A previous article [5] suggested that usage of decision trees, aggregate simulation and delphi methods would dominate micro-simulation and optimisation methods. This is clearly the case—indeed to a dramatic extent in that the only aids actually used by more than 50% of people aware of them (excluding evaluative methods which were not previously discussed): Delphi, Macro Marketing Models, Decision Trees, Sensitivity Analysis and Econometric Models. There is also reasonable evidence to suggest that the problem with many decision aids is clearly lack of usage (and by implication therefore lack of appropriateness) rather than lack of knowledge.

A 'success' rating

We are, of course, interested in levels and growth of usage as well as absolute usage. A weighted index of usage was therefore generated by using the weights indicated in Table 5. It has been assumed that a decision aid which has been tried but discontinued should be penalized more than one which has never been used. The calculated success rating for each aid is shown in Table 6.

The Success Ratings are not significantly different from the usage ratings except perhaps in the case of Heuristic Methods which are clearly used extensively by those who actually use them.

The implications of these ratings are therefore very similar to usage patterns and it is

TABLE 1. DEFINITION OF DECISION TECHNIQUES

Stage 1: IDENTIFY OPTIONS

1.A. *Decision Trees*: A diagramatic technique to present the chronological sequence of both decision options and probabilistic outcomes.

Stage 2: FORECASTING OUTCOMES

2.A. *Econometric Models*: The development of a whole series of relationships often as a result of historical analysis to explain the broad economic response to various policy measures.

2.B. *Marketing Macro Models*: Models derived in a similar manner to econometric models above using broad relationships (viz advertising on sales) derived by both historical analysis and judgement.

2.C. *Game Theory Models*: Models based on specific assumptions about competitive response functions to particular actions, particularly two-company market situations.

2.D. *Sensitivity Analysis*: Often associated with 2.A, 2.B and 2.C above, to develop an understanding of the sensitivity of predicted outcomes to the changes in input assumptions.

2.E. *Marketing Micro Models*: Models derived from micro analytic relationships (viz individual customers response to specific advertising) which are generally used via probability techniques to produce a probability distribution of outcomes.

2.F. *Delphi*: The use of a systematic method to generate a consensus view among a panel of 'experts' on outcomes. This method involves asking individually for outcome forecasts and then providing an opportunity for individuals to modify their forecasts in the light of the average response from others.

Stage 3: SELECT OPTION

3.A. *Discounted Cash Flow*: A financial technique for calculating the internal rate of return of a forecast cash flow stream.

3.B. *Expected Value*: Often combined with 3.A but specifically involving modifying the projected cash flow outcomes by the probability of their occurrence.

3.C. *Risk Free/Risk Trade-Off Return*: Involving modifying the expected return by some means or other to represent the degree of risk (variability) attached to the average outcome.

3.D. *Regret Analysis*: Identifying the opportunity costs of each decision option under all possible outcome situations and identifying the preferred option according to various specific strategies.

Stage 4: MULTI STAGE TECHNIQUES

Really this only involves techniques broadly known as optimisation methods which cover Stage 1 (mainly), Stage 2 (completely), Stage 3 (mainly).

4.A. *Linear Programming*: A formulation of both the variable relationships and also the objectives in a broadly linear and additive fashion.

4.B. *Transportation*: A simplified form of 4.A, often used to derive optimum distribution patterns.

4.C. *Heuristic Methods*: The use of various approximate methods to identify the best solution according to some defined objective.

Vol. 7
1979

TABLE 2.

	Not known	Never used	Tried and usage is now: Discontinued	Infrequent	Average	Very frequent	Current Applications are: Growing	Static	Declining
Decision Trees	0	17	0	11	10	3	7	14	3
Econometric Models	2	17	1	10	6	5	10	10	1
Marketing Macro Models	2	13	1	6	14	5	11	13	1
Game Theory Models	1	28	0	3	5	4	4	6	2
Sensitivity Analysis	3	16	0	7	9	6	5	16	1
Marketing Micro Models	2	22	1	5	11	0	9	7	0
Delphi	3	10	1	14	8	5	5	22	0
DCF	1	6	2	7	10	15	10	19	3
Expected Value	6	14	0	11	9	1	2	18	1
Risk Trade-Off Return	4	17	0	8	9	3	6	14	0
Regret Analysis	6	24	0	5	2	4	4	6	1
Linear Programming	3	29	1	3	2	3	2	6	0
Transportation	8	28	0	2	3	0	0	4	1
Heuristic Methods	8	17	0	1	11	4	0	14	2

TABLE 3.

	Not known	Never used	Tried and usage is now: Discontinued	Infrequent	Average	Very frequent	Current applications are: Growing	Static	Declining
Decision Trees (DT)	0	41	0	46	42	12	29	58	12
Econometric Models (ECM)	5	44	5	45	27	23	45	45	5
Marketing Macro Models (MMA)	5	33	4	23	54	19	42	50	4
Game Theory Models (GTM)	2	70	0	25	42	33	33	50	17
Sensitivity Analysis (SA)	7	42	0	32	41	27	23	73	5
Marketing Micro Models (MMI)	5	56	6	29	65	0	53	41	0
Delphi (DEL)	7	26	4	50	29	18	18	79	0
DCF (DCF)	2	15	6	21	29	44	29	56	9
Expected Value (EV)	15	40	0	52	43	5	10	86	5
Risk Trade-Off Return (RTR)	10	46	0	40	45	15	30	70	0
Regret Analysis (RA)	15	69	0	45	18	36	36	55	9
Linear Programme (LP)	7	76	11	33	22	33	22	67	0
Transportation (TRS)	20	85	0	40	60	0	0	80	20
Heuristic Methods (HM)	20	52	0	6	69	25	0	88	13

Non Known as % of Total
Never Used as % of Known
All other % of Ever Used

particularly noteworthy that Delphi type techniques involving a systematic method for aggregating opinions and providing feedback on other participant opinions are more widely used than is often recognised in the teaching of decision techniques. This is clearly an area for more development work and research and it is interesting that some work is now being done on a more effective framework for assessing the efficiency of various opinion aggregation approaches [1].

Any conclusion about the relative success of a particular technique must be related, of course, to the original definitions used in the questionnaire. There are two particular issues in this case: firstly, the extent to which the definitions are, in themselves, reasonable descriptions of the appropriate techniques, and, secondly, the extent to which the definitions

TABLE 4. NEVER-USED VS. NOT-KNOWN ANALYSIS

	Never-Used (%)	Not-Known (%)
1. DCF*	15	2
2. DEL	26	7
3. MMA	33	5
4. EV*	40	15
5. DT	41	0
6. SA	42	7
7. ECM	44	5
8. RTR*	46	10
9. HM	52	20
10. MMI	56	5
11. RA*	69	15
12. GTM	70	2
13. LP	76	7
14. TRS	85	20

* Evaluative Criterion

TABLE 5. WEIGHTING FACTORS

Not Known	0
Never Used	−1
Discontinued	−2
Infrequent	1
Average	2
Very Frequent	3
Growing	0.5
Static	0
Declining	−0.5

TABLE 6. RANK ORDERING BY SUCCESS RATING

	Rating	Usage Rank
DCF	1.52	1
MMA	0.76	3
DEL	0.74	2
SA	0.61	6
DT	0.51	5
HM	0.46	9
EV	0.43	4
ECM	0.38	7
RTR	0.37	8
MMI	−0.04	10
GTM	−0.10	12
RA	−0.11	11
LP	−0.34	13
TRS	−0.47	14

differ individually in terms of conceptual complexity. On the first issue, it is clearly very difficult to derive universally acceptable definitions but those which were used were seen as relatively uncontroversial. On the second issue, the complexity of the definition for, say, 'Regret Analysis' is clearly much greater than the definition of 'Delphi'. This is partly related to the fact that the conceptual complexity of Regret Analysis is indeed significantly greater than that of the Delphi method and the important issue is therefore whether the complexity of the definition and the language might not mean that individual participants fail to recognise the existence of the more complex techniques. This concern would appear to be of relatively limited general importance since few techniques were indicated as significantly 'not known' and even in these cases at least 80% of participants claimed knowledge. However it is certainly possible that in the specific instance of 'Heuristic Methods' the success rating should be seen as more relevant to the use of well-defined operations research routines compared with the much more general heuristic type of problem solving adopted by managers but rarely described by them as a decision method at all.

CORRELATION MATRIX

It is also possible to generate a correlation matrix to indicate the relationship between heavy uses of one technique and another. The significant elements of the correlation matrix are shown in Table 7. Since there are 91 different elements in the matrix (excluding the diagonal values) it seems most appropriate to concentrate on the correlation factors which are significant at the 1% level or better. A representation of this set of relationships in terms of an association graph is shown in Figure 1.

It is easy to see the three clear groupings:

(i) option generation

(ii) forecasting

(iii) evaluation methods

Similarly, it is clear that in the last two groupings the developed usage of one technique tends to be strongly related to the usage of others, although possibly at very different levels. This is not altogether very surprising when one considers that most techniques in

Vol. 7
1979

TABLE 7. SIGNIFICANT CORRELATION FACTORS

	DT	ECM	MMA	GTM	SA	DEL	DCF	EV	RTR	LP
ECM	0.31									
MMA	0.43	0.49								
GTM	0.45	0.36	0.46							
SA	0.35	0.34	0.60	0.51						
MMI	0.52		0.31							
DEL		−0.35								
DCF						−0.50				
EV							0.41			
RTR								0.52		
LP		0.51	0.36		0.38	−0.42			0.35	
TRS		0.39	0.33	0.46	0.43					0.65
HM		0.31								0.32

Significance levels for $n = 41$
5% $r = 0.30$
1% $r = 0.39$
0.1% $r = 0.48$

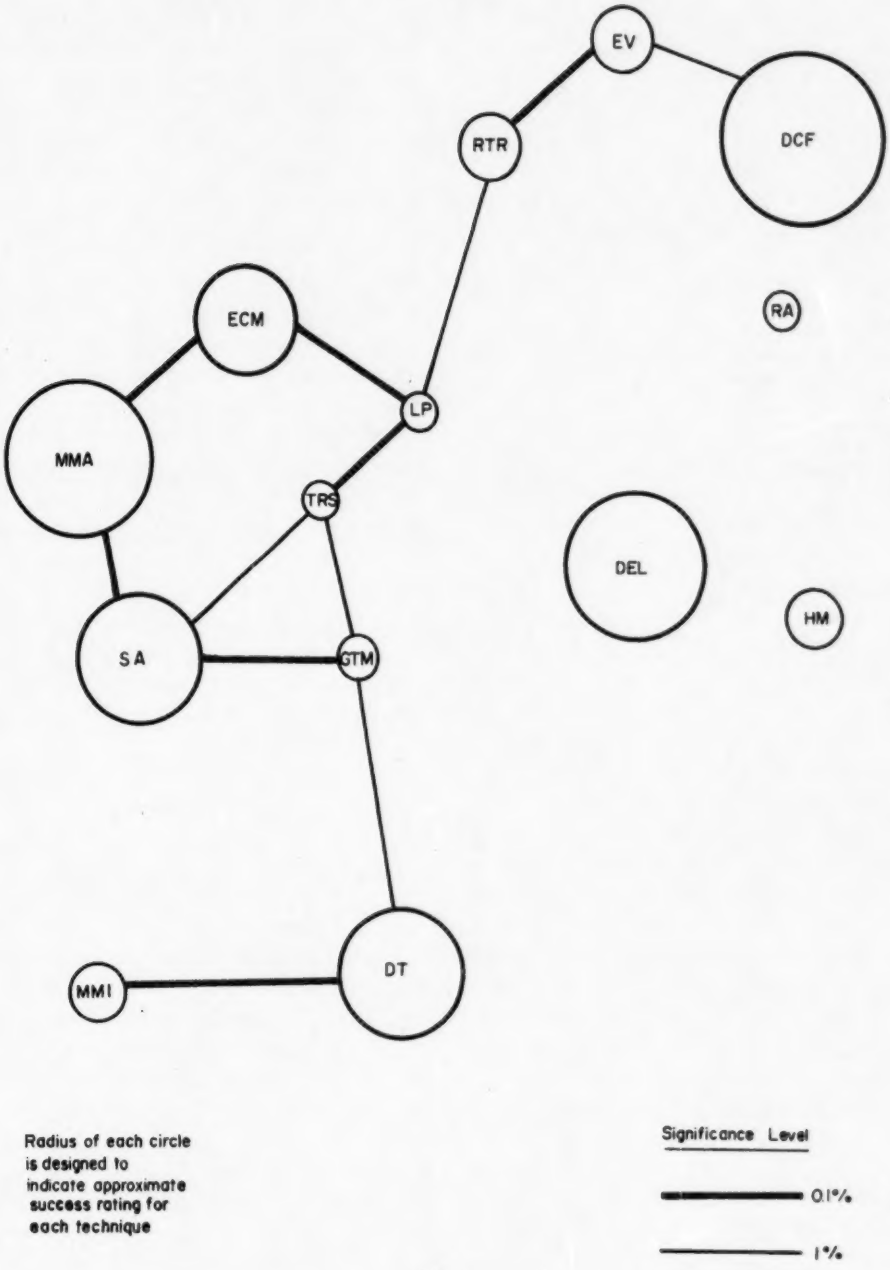


FIG. 1.

one grouping require relatively similar skills and often conceptual approaches. All the linkages are also between what might well be termed analytical techniques (which are of course the substantial majority of the group anyway) and it is interesting that the most subjective method viz. Delphi is negatively associated with this structure of analytical techniques.

It is also interesting to note the lack of strong links between the various groupings and particularly that there is no link in usage terms between Decision Trees and Expected Value evaluation methods. These two are often seen particularly in terms of teaching as almost inseparable and yet it is clear that a more realistic view might be to classify Decision Trees as a method of option generation and decision structuring which is not necessarily related to the means of choosing the appropriate option.

TABLE 8. WEIGHTING FACTORS FOR ORGANISATIONAL QUESTIONS

B1 "In your view, what factors are important in persuading senior management to agree to a certain decision."

Clear Analysis	+1
Analysis of Risk	+1
Subjective Knowledge of Risk	−1
"Feel"	−1
Convincing Presentation	0

Multiplied by importance attached to the relevant factor, viz.

Very Important	=1
Important	=2
Not Important	=3

B2 "Please rank broadly the importance of the following sources of information for you."

Literature	+1
Outside Courses/Consultants	+1
Internal Company Specialists	−1
Own Subordinates	−1

Multiplied by rank for particular factor.

This is consistent with the results of earlier, more restricted empirical research by Longbottom and Wade [3].

Again any conclusion about usage groupings for particular techniques must be related to the respondents' conception of the techniques used and hence to the descriptions offered. To a certain extent the whole layout of Table 1 encourages participants to utilise the major groupings which can be seen in Fig. 1. However the questionnaire did not imply, in general, that the usage relationships should be stronger within these groupings.

The definition did however refer to links between 'Sensitivity Analysis' and 'Econometric Models', 'Marketing Macro Models' and 'Game Theory Models' as well as single links between 'Discounted Cash Flow' and 'Expected Value' and also between 'Linear Programming' and 'Transportation'. In most cases these linkages did appear in the analysis, although the linkage between 'Discounted Cash Flow' and 'Expected Value' was relatively weak and there was no significant linkage between 'Econometric Models' and 'Sensitivity Analysis'. It remains significant that a number of other usage linkages were established whilst others were not.

ORGANISATIONAL FACTORS

The Questionnaire also contained two questions designed to provide an approximate measure of the culture of the organisation and its 'openness'. The relative weightings for the various answers to question B1 and B2 are shown in Table 8. As a result of summing up these weighted responses it was possible to establish crude measures of the 'culture' and 'openness' of the organisation. The culture factor measures the balance between analytical and subjective responses and will tend to zero in the case of a balance, and is negative in the case of an analytical bias, and positive in the case of a subjective bias. Similarly, the openness factor measures the balance between internal and external sources of information and will be negative for an external bias and positive for an internal bias. From the statistical parameters in Table 9 it is clear that most responses indicated a bias towards an analytical culture (in fact only four responses were actually biased towards subjective answers) and it is of course difficult to differentiate between genuine assessment and a degree of bias towards perceived norms particularly for a questionnaire from the London Business School. The openness variable however shows a much more balanced statistical profile.

TABLE 9. STATISTICAL VALUES FOR CULTURE AND OPENNESS VARIABLES

	Mean	σ
Culture	−1.22	1.33
Openness	−0.10	3.47

The significant correlations are shown in Table 10. The results for the culture factor must be treated with considerable caution given the previous comments but it would suggest that an analytical culture encourages the use of game theory, Delphi and Transportation whilst a subjective culture encourages the use of DCF. Similarly, an external bias tends to encourage the use of Risk Trade-Off Return assessment methods but does not have any other significant effects. It is obvious from these results that a simple analysis of the impact of the organisation culture is rather

TABLE 10. SIGNIFICANT CORRELATIONS INVOLVING ORGANISATIONAL FACTORS

	Culture	Openness
GTM	−0.34	—
DEL	−0.33	—
DCF	0.43	—
TRS	−0.26	—
RTR	.	−0.31

Vol. 7
1979

TABLE 11.

	All	Service	Cons/Cons	Cons/Ind	Capital
DCF	1.52	1.00	1.29	1.88	1.0
MMA	0.76	1.30	0.93	0.63	0.25
DEL	0.74	1.70	0.82	0.44	0.62
SA	0.61	0.50	0.50	1.03	−0.50
DT	0.51	1.10	0.61	0.25	0.50
HM	0.46	0.80	0.86	0.31	−0.25
EV	0.43	0.40	0.14	0.72	0.75
ECM	0.38	0.40	0.75	0.16	0
RTR	0.37	−0.20	0.29	0.69	0.75
MMI	−0.04	0.10	−0.07	0.16	−0.88
GTM	−0.10	0.50	−0.18	−0.22	−1.0
RA	−0.11	−0.70	0.04	0.28	−1.0
LP	−0.34	−1.00	0.04	−0.22	−1.0
TRS	−0.47	−1.00	−0.29	−0.34	−1.0
n =	41	5	14	16	4

inadequate. Two points which came across particularly in the detailed interviews need to be considered:

(1) There is often a considerable distinction between how a particular manager makes his own decision as compared to how he 'sells' it within the organisation—a number commented that they used particular analytical techniques to help them make up their own minds but would not have considered revealing the detailed analysis when persuading their supervisors to accept the project.

(2) There are often seen to be other important contingent factors particularly the characteristics of the market place which may well dominate the impact of organisational culture. This is an issue to which the survey results can be addressed although any definition of market characteristics will have to remain rather approximate.

MARKET CHARACTERISTICS

It is possible to segment the total sample into four general market groups:

service industries;

consumer industries;

consumables to industrial customers;

capital goods to industrial customers.

The sample, however, for the first and last categories is very small. The success ratings in each category are shown in Table 11. The rankings of the top five techniques in each category is shown in Table 12. This inevitably indicates some fluctuations and does suggest that in certain sectors, particularly industrial markets, evaluation methods such as Expected Value and Risk Adjusted Return are much better established. In terms of assessing relationships between usage of various techniques in particular market sectors the only category where significant correlations occur partly at least because of sample size is the industrial/consumables category. Significant correlations are shown for this category in Table 13—the basic pattern is not dissimilar to Fig. 1. No significant correlations exist between technique usage and either the culture or openness factor.

IMPLICATIONS

The original thinking behind the survey was that recognised usage of a certain decision aid

TABLE 12. RANKINGS OF FIRST FIVE ONLY

	All	Service	Cons/ Cons	Cons/ Ind	Capital
DCF	1	4	1	1	1
MMA	2	2	2	5	.
DEL	3	1	4	.	4
SA	4	.	.	2	.
DT	5	3	.	.	5
HM	6	5	3	.	.
EV	7	.	.	3	2
ECM	8	.	5	.	.
RTR	9	.	.	4	3

TABLE 13. CONS/IND: SIGNIFICANT CORRELATIONS ($r > 0.59$): 1%

	DT	ECM	MMA	GTM	MMI
GTM		0.63	0.63		
SA			0.81		
MMJ	0.63				
EV					0.60
LP		0.72		0.74	
TRS				0.89	

and research development. It is, however, important to realise that there is also an argument that certain decision aids provide a framework in which to appreciate the advantages and limitations of other more simple approaches—a framework which becomes implicit once it is fully understood. Whilst it is difficult to sustain such an argument in the case of Econometric Models, Risk Adjusted Return and Marketing Micro-Simulation Models, it certainly has some validity in the case of optimisation approaches such as linear programming.

But the situation in which such techniques are not used widely in their own right must generate another series of questions. The value of a technique as an implicit conceptual framework must be compared with alternative methods of structuring or handling the issues. To argue that a certain approach represents the real situation is facile—it is only an approximation to it, surrounded by constraints and limiting conditions. Hence the 'aid to thinking' argument is a value-laden comment—it is better interpreted as an aid to a certain way of thinking which may be beneficial to a certain individual in a specific situation. We have explored this issue much more fully elsewhere, [2] but the basic dilemma can be parodied by the issue of giving effective intuitive thinkers a structured analytic approach which is at best ignored by them and at worst a source of considerable confusion. In this situation it is often more beneficial to build on and develop existing competences than to try and replace them. In this context even relatively simple structured approaches have their dangers—for instance, the simple linear description of option generation, forecasting and evaluation is only an approximation to realities of complex decision making [4].

CONCLUSIONS

1. The overall survey indicates a much wider use of systematic methods for aggregating opinions and providing feedback than is normally recognised. This is clearly a priority area for development work and better theoretical development.

2. The overall survey also indicates the very restricted use of optimisation procedures for marketing situations. This increases the suspicion that such methods are of limited direct value in most marketing situations and that it is often of greater benefit to look at the stages of option generation, forecasting and evaluation as separate but related activities.

3. Forecasting procedures based on aggregate simulation procedures have been reasonably well developed, particularly macro market simulation methods; perhaps inevitably the development of one such method appears to encourage the development of others.

4. Evaluation procedures such as DCF are widely used although it is clear that the usage pattern for these is different from forecasting methods, i.e. they tend to be used extensively or not at all compared with a more normal distribution of usage for forecasting methods. This might well imply that the use of evaluation methods is more a function of organisational procedures than of actual problem characteristics.

5. It is very difficult to evaluate the impact of organisational culture and it has become clear that it is necessary to appreciate the distinction between organisation procedures (perhaps as above) and individual preferences, rather than look solely at the organisational issue.

6. There is no evidence that the relative openness of the organisation in terms of external sources of information has any significant impact on the adoption of specific techniques.

7. Certain changes occur between one market sector and another some of which seem directly related, e.g. as the customer group becomes much smaller the value of aggregate

Vol. 7
1979

market models will become less, and others which probably relate to different industry procedures in assessing options, e.g. in industrial markets there appears to be a greater attempt to incorporate risk adjusted methods.

REFERENCES

1. EINHORN HJ, HOGARTH RM & KLEMPNER E (1977) Quality of group judgement. *Psychol. Bull.* **84**(1), 158–172.
2. LAW PJS & WENSLEY JRC (1979) Marketing teaching. *Eur. J. of Mktg* **13**(1), 52–71.
3. LONGBOTTOM D & WADE G (1973) An investigation into the application of decision analysis in United Kingdom companies. *Omega* **1**(2), 207–215.
4. MINTZBERG H, RAISINGHARE D & THEORET A (1976) The structure of 'unstructured' decision processes. *Admin. Sci. Q.* **21**, 246–275.
5. WENSLEY JRC (1977) Effective Decision Aids in Marketing. *Eur. J. Mktg* **11**(1), 62–71.

ADDRESS FOR CORRESPONDENCE: *Robin Wensley Esq, London Graduate School of Business Studies, Sussex Place, Regent's Park, London NW1 4SA, UK.*

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 25–32
© Pergamon Press Ltd. 1979. Printed in Great Britain
0306-0483/79/0201-0025$02.00/0

The Planning of a Department Store

R FLAVELL

E PENN

GR SALKIN

Imperial College, London

(*Received September* 1978)

A town or village attracts people from its hinterland to purchase goods by virtue of the amount and variety of goods on offer. Similarly, a department store attracts customers away from competitors by means of displays, staff expertise, stock holdings and numerous other parameters. The store is generally free to manipulate these parameters subject to certain restrictions; the manipulation will have a direct affect on the ultimate profitability of the store. This paper describes some current research that is being carried out in conjunction with a department store group. The ultimate aim is to link the corporate planning of the group and each store to detailed socio-economic data describing the population. Some aspects of the problem have been investigated in depth, some are currently being discussed and some have not been considered at all as yet. The latest position will be reported and hopefully some feedback will be engendered.

INTRODUCTION

THIS paper is the initial report of some work undertaken in the UK on behalf of a large department store group. The scope of the research overall is to develop a methodology for the future capital expenditure of the group, and in particular to identify where new stores could be opened, to suggest the size of such stores, to decide whether or not existing stores should be expanded, and so on. At the present time, such expenditure lacks a global view and is thus undertaken piecemeal; the aim of the project is ultimately to present a group-wide view, with all the various interactions this would involve.

However this paper has a far less ambitious objective. It is obviously necessary in this work to understand how an individual department store operates, what are the constraints under which it must act, and what are the various decision parameters under its control. This paper describes the work that has been carried out to develop a planning model of a store; some of the work has been completed, but much is still in the formative stage and it is hoped that this presentation will generate comments and feedback to us.

The paper is divided into a number of parts. Firstly the general problems of controlling a store will be discussed in some detail. This discussion will identify a number of separate aspects whose derivation and measurement will be analysed. Finally the threads will be drawn together with the presentation of various models.

THE STORE MANAGER'S PROBLEM

Certainly in this group, and generally in many such groups, the store manager has considerable autonomy provided that he (or she) is reasonably successful. The manipulation of certain decision parameters is under his *own* control; hence the title of this section.

Superficially his problem is easily stated: For each item in the store, and this may run into

[1] Presented at III Symposium über Operations Research, Universität Mannheim, West Germany, September 1978.

thousands, he has to match supply with demand in such a way as to optimise some objective(s). Supply is very much under his control and as such is a primary decision parameter; unfortunately demand is not fixed and independent of supply nor is it independent of a number of other parameters under the manager's control.

The derivation of a demand function is highly complex. Moving away from the store for the moment, picture a country homogeneously scattered with consumers. Now place in this country a limited number of independent and separate shopping centres; it is not impossible to imaging the consumers gravitating towards the centres as a direct function of their 'attractiveness' and an inverse function of their 'distance'. Briefly completing the analogy, which will be discussed in far greater depth in the next section, it would therefore be possible, in theory at any rate, to compute the potential demand that the consumers will place upon any particular shopping centre. A department store manager in such a centre can do little to influence this potential demand, assuming of course that the size of the store is small in relation to the overall size of the centre. The assumption is likely; department stores are usually (invariably) placed in large shopping centres.

Once the consumer has arrived in the shopping centre, the influence that the manager now has is considerably increased. He is competing using the resources at his disposal with the other shops and stores in the centre for the consumer's attention. If he increases the resources allocated to one particular item, *ceteris paribus*, he must increase the demand in his store for the item. A more detailed discussion of the various resources he can allocate, and the nature of the demand function, is left to a later section.

But the manager's problem has not yet finished. The resources he is allocating are strictly finite and hence the various items are in competition with one another. For example, short of rapidly expanding the store, the total selling floor space is fixed; if more space is allocated to one item, less space has to be allocated to another. The demand for the former may increase, but for the latter will decrease, which could result in a net detrimental effect on the manager's objective. Thus the manager has to decide upon the optimal allocation of these resources, a typical mathematical programming problem, although of course highly complex. The next sections will discuss the construction of the various relationships necessary to solve the manager's problem optimally.

THE DERIVATION POTENTIAL DEMAND

In this section we will examine the measurement of the potential demand for an item (or collection of items) in terms of the success of a shopping centre in attracting consumers from within the surrounding area. For confidentiality, we will identify the particular department store under consideration as D, situated in shopping centre C.

The underlying basis for work in this area is due to Reilly [9], who postulated that competing shopping centres would extract custom from the intervening hinterland in direct proportion to the 'attractiveness' of the centres and in inverse proportion to the square of the physical distance from each centre to the hinterland. Mathematically, one may define

$$p_{st} = (a_t/d_{st}^2)/\sum_{\forall r}(a_r/d_{sr}^2) \qquad (1)$$

where p_{st} is the proportion of customers from area s going to shopping centre t, a_r the 'attractiveness' of centre r and d_{sr} the distance between s and r. The statement is termed the law of retail gravitation after the obvious analogy with the Newtonian law of gravitation. Since Reilly in 1929, there have been a number of suggestions and improvement to this basic model. Perhaps the most recent version is that, after statistically examining some 43 measures, 'attractiveness' has been shown to be highly correlated with the total available floorspace in a shopping centre [8]. Another important modification is to replace the physical distance with a measure of economic distance, i.e. how much would it cost a consumer to travel from s to r, or travelling time.

The actual use made of equation (1) in practice was as follows. First we examined D's delivery and credit files to identify the boundaries of the catchment area. Using the 1971 Census of Population [2], this area was subdivided into 74 wards. This was necessary because a lot of the data we were able to collect was

Vol. 7
1979

only by ward. The next step was to identify the major shopping centres competing over this catchment area. As centre C was situated to the north-east of London, it was felt necessary to subdivide London into identifiable shopping centres, e.g. Oxford Street, Brent Cross. The location and total trading floor-space of each centre was available from a 1971 Census of Distribution, carried out by the Greater London Council [4]. Finally, after identification by planning offices of the precise boundaries of each ward as in 1971, the distance between each ward and each centre was measured in terms of 'main road distance'. It was felt that, as the entire area is relatively wealthy, most people travelled by private car rather than public transport. Hence, a vector containing the proportion of consumers living in ward i shopping in centre C was calculated. The above analysis does ignore the possibility of consumers buying goods near their place of work; it was felt that such purchases tended to be mainly for foodstuffs (which a department store does not generally sell) and that hopefully the total discrepancy was small because we had no way of checking on these sales.

We now have to move from the proportion to the potential demand for each item (or group of items). The 1971 Census of Population [2] identified *pari passu*, the number of households in each ward, whilst two reports by the GLC defined the socio-economic grouping (SEG) of the population within each ward [5, 6]. Finally a recent Family Expenditure Survey [3] describes the income and expenditure distribution, coded by occupation of the head of the household, who may be identified with a particular SEG. Thus, from combining these three sets of statistics, we arrived at D_{iqs}, the demand for item i by SEG q in ward s.

Hence,

$$D_{it}^{*} = \sum_{s} p_{st} \sum_{q} D_{iqs} \quad \text{where } t \equiv C \tag{2}$$

the potential demand for item i in centre C.

As we had no intention of considering each item sold by D individually, we used instead 18 groups of similar items, corresponding to the internal product groups as defined by the store; indeed, throughout this paper, the words 'item' and 'groups of similar items' are synonymous. There were, as one might imagine, considerable difficulties in combining together statistics from different sources and of different ages. We attempted to reconcile figures where ever we could; where we could not, a qualified pragmatism was used.

THE SCARCE RESOURCES

We have so far described how the potential demand for a particular item in a particular shopping centre may be calculated. We will now turn to the next problem, namely a discussion of the scarce resources under the control of the manager and the relationship between these resources and the actual demand in the store.

One could have drawn up a very long list of factors that would obviously influence a consumer into buying in one shop as opposed to another. Such a list would not be very practical; we had discussions with a number of store managers of this particular group and the general consensus was that there were three principal factors, namely floorspace, stock and labour, that would influence demand. Strangely, price was not regarded as such an important factor; on the whole there was little price distinction between shops in the centre unless there was a deliberate attempt to sell off a line.

(a) *Floorspace*

Floorspace is perhaps the most obvious scarce resource. It is obviously limited by the physical size of the store and also, *ceteris paribus*, an increase in the space allocated to a good should increase its relative demand. However in practice it is not quite as simple as that: floorspace is not homogeneous in its appeal. If we adopt as an index profit/sq. ft., then the index decreases as the distance from the ground floor increases. Also the space on each floor may be subdivided; prime space is opposite the main entrance, good space is opposite the side entrances and the mouths of the escalators and lifts, whilst the rest is simply standard space. To increase the space allocated to a good and simultaneously move it to a worse kind of space, need not necessarily increase the demand at all.

This lack of homogeneity presents a major problem. What we would like to do is construct a model that would ultimately compute

the amount of floorspace that should be allocated to a particular item and also where the floorspace should be located. This would require such data as the change in sales as a good is moved around the store: these data are just not available nor are any records kept.

(b) *Stock*

Generally department stores have little storage space, their total stock is usually on display. This of course does encourage the customer if a large, and varied stock is available. But for the manager it does create a problem. As demand is a function of the stock level then one of two situations may occur:

(i) the level of stock needed to generate a certain level of demand is greater than that level of demand, which implies that there will always be unsold items on display. This may create problems of stock turnover and cause low profitability returns per stock item.

(ii) the level of stock needed to generate a certain level of demand is less than that level of demand, which implies that customers will have an unsatisfied demand and sales will be lost. This situation is generally an anathema to managers who primarily see themselves as traders rather than businessmen.

There are further complications, namely that the stock level must be restricted by the floor space available and also by the working capital funds. The store can only afford to hold limited stocks.

(c) *Labour*

In the front-line of any store are the sales staff. Such staff are extremely important as they directly influence the customer. As the level of staffing drops, the waiting time of customers increases and ultimately demand will decrease. But staff is a direct cost to the manager and as such he has to decide upon a suitable level for each item. His task is complicated because many of the staff are specialists and thus not interchangeable.

THE CONSTRUCTION OF A DEMAND FUNCTION

In discussing the key parameters under the control of the manager, we must note that not only do they all directly influence demand, but also that they are all inter-related. One must be careful here to distinguish between demand and sales, for demand may exist independent of sales. It is firstly suggested that the demand function for a particular item is expressed in terms of relative allocations, i.e. the market share λ_i is given by

$$\lambda_i = \frac{D_i}{D_T^*} = f_i\left(\frac{\beta_{ki}}{c_{ki}}\right)$$

where D_i is the demand for item i at D, and β_{ki} and c_{ki} the scarce resources of type k allocated in D and in the entire centre C respectively to item i.

A more precise statement of the form of this function requires a further assumption and we consider two alternative ones.

(a) *Multiplicative demand function*

This is based upon the assumption that if the allocation of any scarce resource is zero, then the resulting demand is also zero, and furthermore that there are decreasing returns to sale, i.e.

$$\lambda_i = \prod_k r_{ki}^{a_{ki}} \; 0 \leqslant a_{ki} \leqslant 1$$

where $r_{ki} \equiv (\beta_{ki}/c_{ki})$. It is felt that the manager would be able to give an indication of the 'relative importance' of each scarce resource in influencing demand; for a multiplicative function, we interpret this statement as an estimate of the relative elasticities of demand. Assuming that resource 1 is the base resource, the relative importance is defined by

$$\gamma_{ki} = \left(\frac{\partial\lambda_i}{\partial r_{ki}} \cdot \frac{r_{ki}}{\lambda_i}\right) \Big/ \left(\frac{\partial\lambda_i}{\partial r_{1i}} \cdot \frac{r_{1i}}{\lambda_i}\right) = \left(\frac{r_{ki}}{r_{1i}}\right)\left(\frac{\partial\lambda/\partial r_{ki}}{\partial\lambda/\partial r_{1i}}\right)$$

or

$$\gamma_{ki} = (a_{ki}/a_{1i})$$

The interpretation is not entirely fortuitous as it does result in an expression that is independent of r_{ki}. The demand function may now be written as

$$\lambda_i = \left(\prod_k r_{ki}^{\gamma_{ki}}\right)^{a_{1i}} \tag{3}$$

Given the status quo (λ_i^0, r_{ki}^0) we can compute

$$a_{1i} = \ln \lambda^0 / \sum_k \gamma_{ki} \ln r_{ki}^0 \tag{4}$$

(b) *Separable demand function*

The underlying assumption here is that demand is generated by each scarce resource

Vol. 7
1979

in isolation, and that the total demand is computed from the addition of the individual demands, i.e.

$$\lambda_i = \sum_k f_{ki}(r_{ki})$$

Notice that this assumption makes the clear distinction between demand and sales, for obviously there can be no sales unless a positive quantity of all the scarce resources is allocated. To produce a logical interpretation of 'relative importance' in this situation is however difficult, and perhaps the easiest direction is to assume that in practice there would be little movement away from the *status quo*, so that a linear demand function would not be unrealistic, i.e.

$$\lambda_i = \lambda_i^0 + \sum_k d_{ki}(r_{ki} - r_{ki}^0)$$

Now, using the same definition of relative importance yields

$$\lambda_i = \lambda_i^0 + d_{1i} r_{1i} \sum_k \gamma_{ki}\left(1 - \frac{r_{ki}^0}{r_{ki}}\right) \quad (5)$$

But there is no simple means of calculating d_{1i} from the *status quo*; the only way is, when the allocation has altered from the original *status quo*, to make a direct substitution.

ALLOCATION OF SCARCE RESOURCES

So far we have outlined the various scarce resources that a manager must allocate and discussed the restrictions and interdependencies that impinge on each one of them. We have also described the construction of a demand function related to the scarce resources and also the potential demand of the entire shopping centre. Before we can proceed further, we need to identify the objective that the manager is attempting to optimise. The evidence for the identification tends to be contradictory: most managers in isolation are primarily concerned with turnover. As stated above, they are principally traders and their general status increases with turnover: furthermore, in the group the pecking order of stores was based upon turnover. On the other hand, taking the group view, store contribution (an approximation to profit) is a much more useful objective and one that the head office is principally concerned with. We adopted the pragmatic approach: as we had been briefed by the head office we used a profitability objective.

As we have seen, it is feasible to formulate this problem as a mathematical program. The profitability is easily calculated as the prices of the goods are fixed, buying costs, costs of floorspace (rent, rates and depreciation) and labour (various hierarchies of sales staff) are all easily obtainable. The different types of scarce resources, their limitations and interrelationships with each other and with the demand function, have been described. So one could solve the allocation problem as described relatively easily although the model might be extremely large. Appendix 1 describes a simple model and discusses appropriate solution methods.

However our first approach for the management of the group was not to use a MP but to develop an heuristic. The heuristic made certain gross assumptions but it had the virtue of being easily and quickly understood. It started by assuming that, just as the division of demand between centres was on the basis of relative floorspace, the potential demand in C could be divided up in proportion to the relative floorspace in C, i.e.

$$D_i = \left(\frac{\beta_{2i}}{c_{2i}}\right) D_i^*$$

where the subscript 2 refers to floorspace. For each item, or rather group of items, various discrete sales levels were identified, S_{ip}, $p = 1, 2, \ldots$. With each sales level was associated four figures: the profit generated by that level of sales (P_{ip}), the space required for those sales (SP_{ip}), the stock required for the sales (ST_{ip}) and finally the level of staffing required to achieve the sales (L_{ip}). Rather than considering every possible combination of scarce resources as the MP would do, the heuristic only looks at certain *a priori* values. One detailed concession was made, namely to divide the floorspace up into different types; in *D* we identified seven grades of space depending upon the floor one was on and one's whereabouts on that floor. This involved, in theory at any rate, producing the above figures for each type of space as well (let the subscript t denote the type of space). Calculate as a profitability index, the profit/sq.ft. $\bar{P}_{ipt} = P_{ipt}/SP_{ipt}$ ($\forall i, p, t$). For each scarce resource, we assume certain known limitations, which are easily

ascertainable from the manager. The detailed heuristic is as follows:

1. Scan the matrix $\bar{P}$ for the highest figures; let this be $\bar{P}_{rst}$. As decreasing returns to scale were assumed, the greatest profitability was generally at the lowest level of sales.

2. Calculate running totals. (i) add SP_{rst} to total floor space of type t so far consumed; (ii) add ST_{rst} to total value of stock held so far; (iii) add L_{rst} to total staff employed so far.

3. Compare the current usage of the scarce resources with the limitations. If any constraint is violated, go to 4. Otherwise go to 5.

4. Identify the accepted item with the lowest profitability. Delete its current level from further consideration and adjust the running totals accordingly. Go to 3.

5. Delete from any further consideration, items (rsu) $\forall u \neq t$. This ensures that the same item will not be spread out on different types of floor.

6. If there exists spare capacity and some items still have to be allocated, go to 1. Otherwise calculate store-wide figures and stop.

Such an heuristic is not optimal, but it does enable the treatment of, e.g. the discreteness of labour, non-convex profit functions, discontinuities in floor space, as well as being extremely rapid to compute. Data collection is a major disadvantage and whilst we have demonstrated its usefulness to the management, the application to an entire store is still in the future.

STOCHASTIC DEMAND MODELS

We now return to the optimising model as this is probably of longer-term interest. One fundamental assumption throughout the work above is that of determinism. Yet, when all the analyses are performed, and the manager has allocated his resources, the final demand is still very much unknown. The sources of this uncertainty are manifold, errors in the potential demand computation, the form of the demand function, a shift in buying patterns, and so on. But any worthwhile model must attempt to take it into account.

How a stochastic demand is to be included is still very largely a matter of discussion. However, for the want of any better information, we made the gross assumption that the demand is distributed with an expected value given by the demand function (D_i) and a fixed exogenous variance (σ_i^2). This implies no increase in the uncertainty as the market share increases. It is supposed that an estimate of this variance may be provided from the historical records of the store, together with the experience of the manager.

In the vernacular, the desired stochastic model is a "here-and-now" one [12] because the resource allocations have to be performed *before* the actual demand can be observed. The class of models may be subdivided further into indirect, corresponding to the active approach of stochastic linear programming [11, 10], and direct. We will concentrate on this latter set, which contains both chance-constrained models and recourse models.

Let us write the deterministic model as

$$\{\max P(\mathbf{x}) \text{ s.t. } \mathbf{f}(\mathbf{x}) \leqslant \tilde{\mathbf{D}};\ \mathbf{x} \in \psi\} \qquad (6)$$

where $p(.)$ is the profit function, $\mathbf{x}$ the set of decision variables, ψ the feasible space defined by the intersection of all the other constraints, $\mathbf{f}(.)$ a vector (of dimension equal to the number of items) representing [sales + expected demand] and finally $\tilde{\mathbf{D}}$ a vector of random variables with zero mean and standard deviation σ. Before we can proceed much further, we have to discuss the effect of this uncertainty on the manager. It implies that there are times when he will have allocated resources to account for a particular level of demand, and that the demand has fallen below this level, thus reducing his profits. The chance-constrained formulation comes to hand immediately; the uncertainty in the model being catered for by $\Pr\{\mathbf{f}(\mathbf{x}) < \tilde{\mathbf{D}}\} > \alpha$. This model restricts the violation of the demand constraint to, at the most, $(1 - \alpha_i)$ proportion of the time, for each item. The certainty equivalence of these constraints is simply

$$f_i(\mathbf{x}) \leqslant F_i^{-1}(1 - \alpha_i)$$

where $F_i^{-1}(.)$ is the inverse cumulative density function of $\tilde{\mathbf{D}}_i$. The uncertainty here is measured by α and it is quite feasible to gener-

Vol. 7
1979

ate a surface $\{p^*(\mathbf{x}), \alpha\}$ as naturally α and p^* are inversely related.

However, such a model does not really get to the heart of the manager's problem. His objective is profit and, at the end of the day, trading risk must be measured in terms of lost profits. Writing the model as

$$\{\max.\ p(\mathbf{x}) - g(\mathbf{y}^+) - h(\mathbf{y}^-) \text{ s.t. } \mathbf{f}(\mathbf{x}) + \mathbf{y}^+ - \mathbf{y}^- = \tilde{\mathbf{D}}; \mathbf{y}^+, \mathbf{y}^- \geqslant \mathbf{0}; \mathbf{x} \epsilon \psi\} \quad (7)$$

where $g(\mathbf{y}^+)$ is a measure of the profits foregone as demand was greater than allowed for, and $h(\mathbf{y}^-)$ a measure of profits lost due to the overallocation of resources. Such a model can be simplified by assuming *a priori*, that the manager would find acceptable a certain minimum profit level (p^*) or a range of market shares, say $[\lambda_L, \lambda_u]$. This would remove the profit function from the objective, which now becomes $\{\min.\ g(\mathbf{y}^+) + h(\mathbf{y}^-)\}$. The model is an example of a programme with simple recourse and as such has been studied in some detail [13, 14]. Unfortunately it is not easy to solve unless some further simplifying assumptions can be made. Appendix 2 contains the details of a simple model and also discusses the implications of other assumptions.

SUMMARY

It is very common, in the UK at any rate, to find that managers of department stores do not regard themselves as businessmen at all but more as traders. This has obviously got an historical basis. In this paper, we have discussed some of the problems of running a store and have attempted to display a scientific approach to these problems. Like any business, at the heart of a store is the economic relationship between supply and demand. We have constructed a demand function for the various items in the store and examined how this demand will vary as the manager allocates the scarce resources under his control. We have outlined various models that may be used as the basis of any structural planning process and suggested algorithms that may be implemented. Uncertainty plays a large part in the manager's problems and we have proposed methods of manipulating it so as to reflect its true role.

We have not attempted to answer the fundamental question: what makes a department store different from a collection of individual shops. We could put forward suggestions related to synergy, ambience and so on, but ultimately it is a question we are unable to answer. There are other facets that have not been discussed. In retailing there is the theory of complementarity—namely, that goods that tend to be sold together or to the same group of customers are put in close physical proximity. This can have a major impact on the layout of a store. One technique we think might be applicable here is graph theory, with weights on the arcs denoting the degree of complementarity and trying to identify spanning trees or restricted degree. This would certainly be a very difficult problem to handle using programming techniques.

The demand pattern for goods may be highly seasonal, e.g. toys, garden tools, and yet all of the models so far presented do not possess a time dimension. One answer might be to produce a multi-period model, but perhaps a more realistic one is to run the static model at, say, monthly intervals with appropriate (and largely subjective) modifications to the demand function. Data are just not available to build up monthly demand functions in the manner suggested in the main text. Finally, to reiterate the limitations we laid down at the beginning of the paper, there is deliberately no discussion concerning capital expenditure and the more major *group* task of assessing and controlling a large number of stores. These topics will be the subject of a future paper.

REFERENCES

1. Baumol WJ & Bushnell RC (1967). Error produced by linearisation in mathematical programming. *Econometrica* **35**(3–4), 447–471.
2. *Census of Population* (1971). H.M.S.O., London.
3. Family Expenditure Survey (1975) Department of Employment, H.M.S.O., London.
4. Greater London Council (1975) Note on 1971 Census of Distribution for London and the South-East.
5. Greater London Council (1972) Demographic, social and economic indices for wards in Greater London. Research Report No 15.
6. Greater London Council (1976) Social and economic indices for wards in Greater London. Research Report 20(2).
7. Himmelblau DM (1972) *Applied Non-Linear Programming*. McGraw-Hill, New York, Chap. 6.
8. Kilsby DJE, Tulip JS & Bristow MR (1972) The attractiveness of shopping centres. Occasional Paper No 12, Centre for Urban and Regional Research, University of Manchester, UK.
9. Reilly WJ (1931) *The Law of Retail Gravitation*. New York.

10. SENGUPTA JK, TINTNER G & MILLHAM C (1963) On some theorems in stochastic linear programming with application. *Mgmt Sci.* **10**, 143–159.
11. TINTNER G (1960) A note on stochastic linear programming. *Econometrica* **28**, 490–495.
12. VAJDA S (1972) *Probabilistic Programming*. Academic Press, New York.
13. WETS RJB (1966) Programming under uncertainty: the complete problem. *Z. Wahrs Revw. Geb.* **4**, 316–339.
14. ZIEMBA WT (1970) Computational algorithms for convex stochastic programs with simple recourse. *Opns Res.* **18**(3), 414–431.

ADDRESS FOR CORRESPONDENCE: *Dr R Flavell, Department of Management Science, Imperial College, Exhibition Road, London SW7 2BX, UK.*

Appendix 1: A deterministic model

The aim of this appendix is to demonstrate how a deterministic mathematical programming model of a store may be constructed, to build a simple model and to suggest various appropriate solution strategies.

Concentrating firstly on an individual item, say the ith one, the supply and demand constraints must be constructed. The sales of this item, S_i, are limited by:

the stock of goods, β_{1i}
the floorspace allocated, β_{2i}
the labour force available, β_{3i} and, of course, the demand, D_i.

The precise nature of the limitations depends upon the required detail, as discussed in the main text, but the simplest assumptions of homogeneous floor space and sales staff might produce constraints of the form

$$S_i \leq \beta_{1i}$$
$$y_i\beta_{1i} \leq \beta_{2i}$$
$$S_i \leq \theta\beta_{3i}$$
$$S_i \leq D_i$$

where y_i is the floorspace consumed per item and θ_i the selling capacity per member of staff. Fractional staff are deliberately implied, equivalent perhaps to staff working in more than one section of the store or the employment of part-time staff.

The demand, being a function of the various β's, could be as described either in equations (3) and (4) or as in (5), i.e. multiplicative or linear. If the multiplicative function is adopted, then an obvious strategy is to employ geometric programming: this in addition would be able to handle easily conditional demands between items. Alternatively, because D_i is monotonic in all the β's, it is possible that a simple application of an iterative linear approximation technique (see [7] for further details) might suffice [1]. Finally, if only small variations in the demand around the status quo are considered to be likely or permissible, use of a first order approximation of a Taylor series, together with the fact that the second differentials are all non-negative would result in a constraint of the form

$$D_i \leq D_i^0\left[1 + \sum_k \left(\frac{\beta_{ki}}{\beta_{ki}^0} - 1\right)\gamma_{ki}a_i\right]$$

The profit contribution of this item consists of the gross operating profit less various charges for excess stock holding, labour and, possibly, the use of floorspace and other facilities. In the simple model, this may be written as:

$$(p_i - c_i)S_i - rc_i(\beta_{1i} - S_i) - d\beta_{3i} - e\beta_{2i}$$

where p_i and c_i are the price and cost of the item, r the interest charge on stock holding, d the labour rate and e the charge per unit of floorspace.

Looking at the overall picture, there are restrictions on the various decision parameters relating to, for example, the physical size of the store, the size of the labour force and the available working capital. Simple descriptions of these constraints would be:

$$\sum_i c_i\beta_{1i} \leq M \quad \text{—current asset constraint}$$
$$\sum_i \beta_{2i} \leq N \quad \text{—total floorspace}$$
$$\sum_i \beta_{3i} \leq \Xi \quad \text{—total level of staff}$$

but of course they could be far more complex. Examination of the shadow prices attached to these constraints would be of considerable interest, both to the manager and the head office, as they might form the basis for inter-store efficiency measures.

Appendix 2: A stochastic demand model

We demonstrated in the main text that one approach to handling a stochastic demand might result in the following model:

$$\{\max.\ p(\mathbf{x}) - g(\mathbf{y}^+) - h(\mathbf{y}^-)\ \text{s.t.}\ \mathbf{f}(\mathbf{x}) + \mathbf{y}^+ - \mathbf{y}^- = \tilde{\mathbf{D}};\ \mathbf{y}^+, \mathbf{y}^- \geq 0; \mathbf{x}\epsilon\psi\}. \quad (7)$$

If we make the assumption that the penalty functions are both linear, e.g. $h(\mathbf{y}^-) = \Sigma_i h_i y_i^-$ *and* that we are concerned about the *expected* violation of the demand constraint, so that $g(\mathbf{y}^+) = 0$, then the objective may be written as:

$$\max.\ p(\mathbf{x}) - \sum_i h_i \int [f_i(\mathbf{x}) - \tilde{D}_i]\, dF(\tilde{D}_i)$$

integrating over the range

$$f_i(\mathbf{x}) \geq \tilde{D}_i$$

For an appropriate distribution, e.g. uniform, the objective becomes

$$\max.\ p(\mathbf{x}) - \tfrac{1}{4}\sum_i \left(\frac{h_i}{\hat{D}_i}\right)[f_i(\mathbf{x}) + \hat{D}_i]^2$$

where $\tilde{D}_i \sim R[-\hat{D}_i, +\hat{D}_i]$. This may be easily solved using either a quadratic algorithm (if the LP approximation for $f(\mathbf{x})$ is suitably valid) or some non-linear algorithm.

Another approach, because the integral is in the objective function, and because the integral is monotonic in $f_i(x)$, we could use again a version of linear approximation programming.

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 33–41
© Pergamon Press Ltd. 1979. Printed in Great Britain
0306-0483/79/0201-0033$02.00/0

Futures Literature and Futures Forecasting—A Critical Review

BHP RIVETT

University of Sussex, UK

(*Received July* 1978)

This paper presents a personal view of futures forecasting. Beginning with an overall view of the literature in the field with particular attention to the more significant contributions, methodology is then considered. Various general techniques of very long term forecasting are reviewed, followed by consideration of the particular fields of economic, social, technological, energy and politics. The social factor is seen to be all-pervading. A synthesis of futures forecasts is then attempted with particular reference to the British case. Finally, the significance of such forecasting approaches for the OR scientist is considered.

INTRODUCTION

IN RECENT years an important literature has sprung up, almost unnoticed by the O.R. practitioner. This literature is concerned with forecasting what the world will be like at times beyond the reach of mere extrapolation of present trends. It is, therefore, in some sense of a poetic nature and alien in its approach to the physically trained scientist who likes to observe, to measure and to model. But we are forced to forecast beyond the limits of extrapolation where we are dealing with decisions which will have consequences over long time periods ahead; for just as prudence advises the navigator of a ship to be able to see at least as far ahead as his stopping distance, so must our models incorporate forecasts at least as far as our decisions will commit us.

Since knowledge in this field, therefore, inevitably incorporates a large subjective element, so will this paper be subjective and be at the mercy of the emotions of the author. The points of view are those of a quantitatively oriented scientist with no track record in the futures field. For this reason the paper will, from time to time, use the personal pronoun in order to emphasise its personal nature.

This paper is based solely on a literature review. In the course of the research on which it is based, I was privy to a number of highly confidential discussions with senior personnel in industry and government but nothing of what was said is quoted here although it must, of course, have influenced my views.

The first questions in such a survey concern the constraints of geography and time. It is tempting to take a world view of the next hundred years or so, but the more modest boundaries set here are of the UK to the year 2000.

Secondly it is borne in on one viewing these issues from Britain, that this country is a small one in which its people are mixed up in each other and densely packed. (If the whole of the present population of the world was put within the USA, the density of population would only be twice that of our's today in Britain.)

The interlocking nature of the UK of the present and the immediate future has meant that the only view one can take is systemic. Hence, although the first division of the literature was into the natural tetrarchies of economics, social, technological and political, such a split could not retain credibility and, as will be seen, it contains one pervasive ubiquitous member which cannot be isolated from the other three.

The literature is wide and in the course of this survey nearly a million words have been read. The quality is unnervingly variable. In the more quantitative areas, such as economics, it is possible to confirm or refute conclusions

drawn by analysts because they are based on published data and methods of analysis which are open and declared. In other areas analysis disappears and one listens to the words of those who hear voices and see visions and it is, of course, more difficult to discuss or refute a vision than a quantitative analysis.

One of the variabilities which exist in this literature is the actual time horizon which has been taken by writers in various fields. Long-term might be taken as referring to time horizons at which the problems which will be encountered cannot be deduced by mere extrapolation from the present. This time scale will vary with the subject. For example, in population and energy, concern is focussed upon periods well into the next century. For agriculture, long term forecasting consists of anything more than ten years ahead. Consequently it is difficult to apply rigidly the twenty year horizon to which we have referred.

Public interest in the relation between different aspects of the future scene, population, resources, pollution, seems to have been stimulated by the publication and discussion of "Blueprint for Survival" [6], and popular accounts of early mathematical modelling such as "World Dynamics" and "Limits to Growth" by Forrester, Meadows and others at M.I.T. [3, 11]. More recently this interest has been reinforced by the second report to the Club of Rome—"Mankind at the Turning Point". The assumptions, methods [12] and conclusions of some of these earlier studies were discussed in particular in two works by the World Bank [13] and by the Science Policy Research Unit at the University of Sussex [1, 5]. At the same time there has been a growing awareness of the increasing interdependence of nations which has developed through a number of international gatherings including the World Population Conference in Bucharest (1974), the United Nations special session on raw materials and development (1974) and the World Food Conference in Rome (1974).

In the UK interest was stimulated and extended by the work of the Systems Analysis Research Unit in the Department of the Environment, reviews of the world-wide long term availability of resources, particularly food, and the effects of market pressures and pollution control on the United Kingdom. The Chief Scientific Adviser to the Government published a report in Nature [2] which was based on the first findings of some of this work.

METHODOLOGY

Basic to futures study is the provision of forecasts, sometimes quantitative, sometimes qualitative, but the literature in general fails to make an obvious distinction between three kinds of forecast. There is first of all the forecast which is self-fulfilling. For example, if a commodity buyer anticipates a price increase, then his reaction will tend to encourage a price increase. The second category is the self-negating forecast. If it is stated that in the future the streets of Brighton will be unsafe to walk on at noonday, the reaction of the community will be such as to negate that forecast. The final category is the neutral forecast which has no effect (such as a forecast that in five years time Millwall will win the first division football championship in Britain). In considering forecasts it is important to note which category they are in, and, even more important, those occasions where a series of forecasts derived separately lead to consequential actions which reinforce each other and enhance the likelihood of what is forecast.

It is apparent from the literature that the most secure forecasts are based on a number of separate approaches. Good forecasting makes use of quantitative data, local, national and international, but where possible these are moderated by the use of judgment and intuition. Economic indicators display trends in business cycles but their usefulness extends only over 6 to 8 months and are best used as descriptors of the present state. Periodic surveys of the views of businessmen are also of only short term value as forecasters of up to a year ahead. The standard tool of economic forecasting is regression analysis as a way of identifying trends, arithmetic, exponential, logistic and logarithmic and such methods may extend our horizon of vision to three or four years.

In these cases one is dealing with situations where variables can be assumed to operate independently of each other. If there is an understanding of the way economic variables affect each other then we can use an input–output table within and between industries and within and between geographical areas. But this has

Vol. 7
1979

limited use in forecasting largely due to the delay in producing the tables. For example the UK input–output analysis tables of 1973 were based on 1968 data.

Simulation methods are obviously attractive to the O.R. Scientist and are well described in a number of sources. It is suggested they have validity for up to 15 years ahead but for the longer term forecasts which are our present concern, we need to break out of the short term constraints of all these methods. The basic approach is to form scenarios which give a qualitative or quantitative description of the world at a particular future time. The scenario describes scene by scene one or more mental images of the future and for the method to be informative it is important that these images be detailed, self-consistent and real.

Early on in every scenario exercise two crucial decisions must be made:

(a) The timescale.

(b) The number of options.

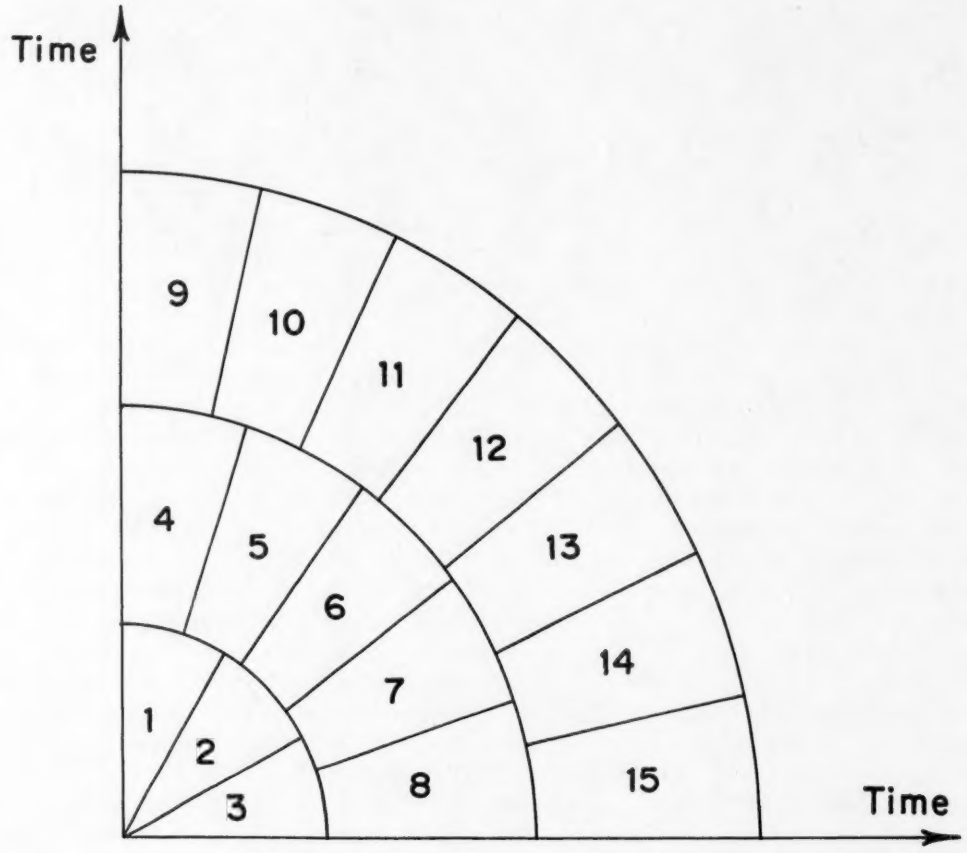
In time, periods up to 15 years are, as stated, better dealt with by simulation and although the number of options may be infinite, the minimum number seems to be three.

The Delphi technique is well known outside O.R. although many practising O.R. Scientists do not appear to have used it. The method evolves forecasts from a group of people but avoids face to face discussion and confrontation in which the strongest personality, or loudest mouth, may dominate. Forecasts are made anonymously, are exchanged and criticised with (one hopes) a gradual movement towards a consensus view. In practice Delphi coordinators might submit questionnaires to the individuals of a group and report back to them (individually) in the form of edited reports and fresh information. From such a series of questions and answers the group and the coordinators move to a series of forecasts, opinions and scenarios. The opinions of the participators can be focussed upon the provision of a series of scenarios stretching into the future.

Another method of producing future scenarios is to reverse the process of Delphi. In this one writes a description of the world as it might be in the year 2000 and then from the perspective of the year 2000 the history of the previous 25 years is written. In this history, which would take us from the known of 1975 to the suggested world of 2000, one can identify critical points of change in the world which lead to the culmination of the particular history described for the year 2000. By analysing such 'historical' documents one gains an understanding of the events which might have to occur for particular scenarios to obtain in reality. Attention can then be focussed on the likelihood of these events occurring and hence on the likelihood of the particular scenario being evolved.

We now move on to the essentially quantitative methods. All of these revolve in some way around the methods of simulation which have been mentioned earlier. We have referred to the interconnectivity of the matters which one is dealing with in future developments and these interconnectivities, in the quantitative approaches, are dealt with by means of correlations or stated relationships. For example, in using a cross impact analysis one is concerned with stating the relationships between various events which might occur and the subsequent events which are affected by these causes. Sometimes these relationships are deterministic in nature, rather akin to the input–ouput tables which we discussed earlier. In other cases the occurrence of an event will affect following events in a probabilistic form. In cross impact analysis one states these probabilities and uses a simulation in order to produce various alternative future scenarios. In practice the method is fiendishly complex and complicated because of the combinatorial problems involved. There is a convenient shortening of the cross impact method by what is called the KSIM approach developed by Kane of the University of British Columbia, Vancouver [7].

Divergence mapping is a pictorial approach for moving from the present into the future. It rests upon the assumption that we are most sure about the present situation and that we are going to move into the future along a series of alternative snapshots, all of which are related to each other by time. The diagram which follows shows the way in which this would proceed. Diagrammatically the areas on the map are different scenarios where time is taken as moving out from the central origin. One can only move from one scenario to



another as a function of time, and these scenarios are adjacent in the mapping. This is a way of relating a series of scenarios and seeing the extent to which the movement over time is plausible. The method is attractive, as indeed is any pictorial method of dealing with these complex forecasting problems.

In summary, from my understanding of the literature it would appear to me that the main method of producing forecasts in the medium term is by the qualitative discursive methods referred to in the earlier part of this section rather than by the formal mathematical methods.

ECONOMIC FORECASTS

Economic forecasts of the future are prolific. We shall merely restrict ourselves to a summary of one in particular. In doing so, it is useful to remember that this most detailed and open long term economic forecast for the UK by Leicester [9], which was published in 1972, has little or no reference to those magical words "North Sea Oil". Such is the way in which new features can arrive quickly on the scene.

The methods used in Leicester's forecast are at the quantitative end of the above summary of techniques but include as conclusions

(1) By 2000 the level of economic activity will be between two and three times current levels (corresponding to growth rate of at least 3.4%).

(2) The largest share of growth will go into investment as opposed to consumption and the next largest into social services.

(3) Per capita consumption levels will double—people will live in better houses, with 1.5 cars per household, travelling 3 times as much; their health will be three times as well cared for (in expenditure terms) and individuals will be better educated.

A social implication of this growth is the

extent to which the gap between rich and poor nations will be allowed to grow. Martel [10] points out that over the last 100 years the gap between the ratio of the richest 10% and the poorest 10% of nations has widened. At the beginning of the industrial revolution it was 5:1, today it is about 200:1. This gap might be changed by unionisation and consequent high wages increasing the attraction of less developed countries in providing pools of cheap labour.

SOCIAL FORECASTING

In this area we are dealing with an ephemeral but pervasive effect. Ephemeral because of the rapidity with which social factors ebb and flow. The rapid increase in the role of women in society will have profound effects but is comparatively fresh. The car is now a social necessity but even 25 years ago the standard design of a house did not incorporate a garage. Central heating is now as essential as good sanitation, but even 15 years ago it was an optional extra in a new house. Entertainments are ephemeral; no longer is cock fighting (or even boxing) a popular spectator sport. Speedway languishes and tenpin bowling had a life (in the UK) of about 10 years. Twenty years ago computers were glamorous and exciting, now Woolworths sell small computers. We seem to be moving towards an earlier retirement age and this, combined with delay of entering employment caused by education and increased life expectancy, will mean that less than half our lives will be spent working for a living and so leisure industries will increase in importance.

But there are social constants. One is geography—not only the natural boundaries of rivers and hills, but the astonishing permanence of the man-made urban environment. The layout of towns and cities changes very slowly. Even the north and east boundaries of Hyde Park follow Roman roads and have existed for 2000 years.

The sizes of communities (in relation to each other) have changed little over the years. In the USA over the last 30 years the proportion of the population changing residences has remained almost constant at 20%. In this country data for Dorset (for example) show that between 1847 and 1936 the distance separating marriage partners did not exceed 19.5 km in 87% of cases and for the country as a whole in the decade 1953 to 1963 the 87% figure was constant for distances up to 16 km.

TECHNOLOGICAL FORECASTS

It is always difficult to forecast the direction of scientific development. Scientists in the public domain (for example universities) are jealous of their freedom of research while those in the private sector may be bound by secrecy. Nevertheless we have less excuse for being taken by surprise by technological change since it casts before it a long shadow of basic research, applied research and development.

The literature reveals a number of interesting points. The thrust of inventions in electronics and computing is probably enough well known to the readers of this journal for any account here to be repetitious. Mathematical Modelling of the rate at which a new product substitutes for an old one shows interesting results.

There are dangers, not always recognised in this kind of forecasting, of failing to recognise a re-entrant technology. Thirty years ago the gas industry was written off by all discerning thinking people. The advent of high speed British Rail virtually destroyed the burgeoning short haul air sevices between London and Manchester and resuscitated rail travel. We now see the rapid resurrection of radio and are even hearing cries for the return of the tramcar to our city streets.

POLITICAL FACTORS

The first sentence of a paper by Rose is: "A paper about the role of forecasting in political science cannot be a review of the literarure, because there is no literature". As Rose states, the basic difficulty seems to be intellectual. If an economist talks of economic change the meaning is usually clear and capable of measurement. When a technologist talks of change what he is implying is visible physically. When a sociologist talks of social change the meaning is less clear and less readily subject to quantified measurement but nevertheless ought to refer to a specified model of the social system. But when political scientists talk of political change it can mean anything (for example, a change in the party controlling a

local authority), everything (a totalitarian style revolution), or nothing.

The most important obstacle is the infrequency of observational phenomena. Only four people have led the Labour Party in the last 40 years and only eight have led the Conservatives. The circumstances in which they come to power may be fortuitous (the death of GAITSKILL and the illness of MACMILLAN). As Rose points out, where quantified data are provided they achieve their objective by accumulating quantities of data that are economic or social rather than political. For example, the book, "Political Change in Britain" by Butler and Stokes, published in 1969, is misleadingly titled. To analyse the subject it purports to study, the authors would have had to deal with the politics of three decades, three generations or three centuries, rather than, as was the case, with three sample surveys undertaken in a time span of less than thirty-six months.

The last twenty years have seen the emergence of a science of political forecasting. We have seen the growth of polling and studies of the reasons why electoral campaigns have been fought the way they were. However, most of these studies are centred specifically around a General Election. We do not have continuing studies showing the way in which political views are changing within the population, and the reasons for those changes. We have a great deal of detailed study of General Elections themselves but in looking at a continuous process of political change over a period of years, one is really faced with a series of detailed snapshots at election time with very, very little picture of what is happening in between elections.

Most of our forecasts of political change stem then not from quantitative analysis as such, but from forecasts derived as a result of discussions between experts and industrial managers.

The political forecast here is of a continuing increase in power of the E.E.C. It seems to be conceded that such changes are long term and will continue. What is less sure is the extent to which the present moves towards regionalisation in Britain are as definite as the moves towards centralisation in Europe. It may be that the social changes which we have seen come about by means of the devolution movements, will be ephemeral. However, one's impression from the literature and from the reports which I have read, is that we are, indeed, faced with the continuing move towards local organisation.

In the final section of this paper we draw a parallel between the number of ways in which two tendencies, towards corporate decision making and towards individual decision making, seem to be progressing simultaneously. On this basis it seem fair to suppose that the movements towards regionalisation and devolution in the United Kingdom, will continue.

The incoherent nature of this section is partly a reflection of the literature on which it is based. There is indeed no such literature and there seems no science of political forecasting as such.

ENERGY FORECASTS

This is a literature (even a cottage industry) all of its own and has a major share of the market in futures literature. It would be naive to summarize in a few lines, but certain points emerge. This is, first of all not only the largest sector of the forecasting industry but also that in which the largest errors have been made. In 1866 Professor WS JEVONS forecast that within 100 years the demand for British coal would be 2607 million tons a year (in terms of coal *equivalent* in 1966 energy consumption in the UK was 298 million tons). In 1966 UK coal production was 176 million tons. The Robinson commission of OECD (1960) stated that no persistent shortage of primary energy was likely by 1975 (correct) but added the incorrect rider that there was no real need to create new sources of energy [4].

What reliance can be placed on long term energy forecasts? Many forecasting exercises are little more than statistical extrapolations which rely on the assumption that past relationships will continue, rather than being based on an appreciation of the possibilities for economic, social and technical change. Nothing seems to be known about long term energy price elasticities or substitution forecasts. The basic data seem of poor quality—reliable time series for energy consumption were rarely published before 1950 and little can be found about relative demands for different types of power. Even the basic assumption that the correlation between economic growth and

Vol. 7
1979

energy requirements will continue and may have to be examined; for example energy conservation may affect the relationship between these two factors.

The basic reserves of fossil fuels and uranium deposits are not known with any precision. Whereas the literature tends to treat energy as a finite resource within a statistical analysis framework, the size and composition of 'reserves' keep changing; for example, wood, peat, coal and animal power have been steadily replaced by petroleum, which may itself be complemented by new energy sources.

It would be unrealistic to expect technical progress now to come to a halt and too complacent to rely on it and market forces to provide timely and socially desirable solutions to the energy problems of the future. There do seem to be major uncertainties and this leads to two sharply different schools of thought—the 'apocalyptic' and the 'Lord will provide' school. The analyses of apocalyptic school point to an exhaustion of energy reserves, accompanied by strife and wars as nations fight over dwindling supplies, early next century. (The lack of firm knowledge of resources renders dating difficult). The other school points to the historical evidence of replacement of one form of energy by another (muscle, wood, dung, coal, gas, oil, nuclear) and believes that economic forces and substitutions will lead to further supplies which will be sufficient especially when energy conservation becomes socially desirable.

THE PERVASIVE SOCIAL FACTOR

We have referred above to the ephemeral nature of social concerns. These also impregnate the technical, economic and political and inevitably impose a systemic view of the future.

Two examples of this may illustrate the point. Few could have forseen that the social desirability of the first motor cars produced nearly one hundred years ago would lead to the development of machines which (in the USA) would kill 50,000 people a year, pollute the air of major cities, dominate the economics of the world, radically affect the political scene over a large area of the world and lead (perhaps) to wars.

Another example of this is from the United Kindom. A consequence of increasing social aspirations is that more people go into further education. Amongst these are the skilled and intelligent who previously filled jobs such as gas fitters. The resulting shortage of these skilled men is that gas apparatus used in the home might have to be redesigned so that unskilled labour (including the householder) may deal with it, as is the case with simple electrical connections, for example. This example shows links between social, economic and technological factors. Clearly such social attitudes affect economics. Propensity to save is a social affair, as are the attitudes to energy saving (smaller cars) and attitudes to public transport.

A SYNTHESIS

All those concerned with forecasting (including the writer and the readers of this paper) are statistical freaks and are not average people. They are part of the articulate middle class, dominantly home counties based, and their projections of the future stem from their culture and attitudes. This not only affects the answers to the questions posed but also affects the questions themselves. In so far as the act of forecasting is not neutral in its effect, then this particular freak class will have a leverage effect on the future.

Within the wide range of literature studied a number of strands emerge, sometimes explicitly but more often than not implicitly, so far as UK 2000 is concerned.

(1) We shall not be involved in a major war (this possibility is hardly mentioned in the million words read).

(2) Economic Growth will continue to be a main objective. With a 3% compound growth rate, this implies a doubling of living standards by the year 2000.

(3) Service Industries will grow, particularly health and education.

(4) Geographical layout of towns and cities will change slowly, with gradual migration within the UK.

(5) The gap between rich and poor nations will not narrow.

(6) Women will take up more senior posts formerly occupied by men. The increase in the number of university graduates will lead to an upgrading of the skills of many jobs.

(7) Most important of all—there will be two competing social forces acting on society—a conflict between what makes for efficiency in technoeconomic terms and what makes for social well-being. This may express itself, crudely, in conflict between corporate and individual solutions.

The corporate pressures include:

(a) Political control over more aspects of national life from outside the UK—Brussels or a European Parliament.

(b) Large multi-national companies dealing at supra national level with other multinationals.

(c) More and more decisions of a company being taken at the centre (following the ability to centralise data flow) and a consequent loss of power from regions to the centre.

(d) The political movements towards the corporate state.

Examples of these are or might be:

(i) Direct elections to a European Parliament.

(ii) Laws to deal with multi-national activities.

(iii) Loss of power and status of executives in the provinces compared with London.

(iv) The emergence of trade unions as dominant partners in national planning.

(v) Large shareholders increasing their blocks of shares at the expense of the individual shareholder.

(vi) Pressure on the consumer to use public rather than private transport.

(vii) The raising of funds and of social engineering by means of pay as you earn income tax.

(viii) National Health Service, larger universities.

(ix) Municipalities offering housing, e.g. council houses, for rent.

The contrary movements include:

(a) Devolution.

(b) Do-it-yourself technological developments.

(c) Participation in decision making.

Examples of these are:

(i) The relative sudden emergence after years of ineffectiveness, of movements for devolution in Scotland and Wales.

(ii) The movement away from laundries and laundromats to home laundries. Dishwashing machines, frozen meals (replacing, as a convenience, restaurants) mini calculators and mini computers.

(iii) Attempts at industrial democracy. (The student participation movement started in 1968. The same students are now in their thirties and moving up the management ladder).

(iv) The popularity of council house sales. Individual home ownership. The Open University (the apotheosis of 'do-it-yourself' education). The success of private health care and private education.

(v) The emergence of shop floor power in unions, the 'wild cat' strike.

(vi) The increasing independence of women with a woman being a woman in her own right and not simply ex-officio as mother of a family.

(vii) The continuing increase of car ownership, particularly the second car.

Not all these battles will go the same way, in fact corporate solutions to some may lead to compensatory individual solutions in others and vice versa. But clearly we need to create some form of leading indicators which will

Vol. 7
1979

show how things are going. These indicators exist in economics and are greatly needed in the social field.

But what are the implications for the O.R. scientist? Firstly we should become acquainted with the techniques, results and literature, of futures research. Indeed, more O.R. groups and university departments should actually be doing it. These environmental factors of the economy and technology are macro forces and we should know them, understand them and incorporate them in our modelling. Otherwise we are in danger of being reduced to the level of usefulness of bar room stewards of the Titanic for whom even being promoted to serving in the first class bar (which is the ambition of so many) will not significantly alter our eventual destiny.

ACKNOWLEDGMENTS

I would like to express to all those who spared time to assist me in this work by discussing their own and their organisation's views of the future. For reasons of confidence they have to remain anonymous but they will know from this paper the extent of my indebtedness.

REFERENCES

1. COLE HSD (1973) *Thinking about the Future: a Critique of the Limits to Growth.* Edited for the Science Policy Research Unit, Sussex University. Chatto and Windus for Sussex University Press.
2. COTTRELL A (1973) Problems of predicting future world trends. *Nature* **245**(5424), 280–281.
3. FORRESTER JW (1971) *World Dynamics.* Wright Allen Press, Cambridge, MS, USA.
4. FREEMAN C, JAHODA M & MILES I (1976) *Forecasting in Political Science in Progress and Problems in Social Forecasting.* Social Science Research Council, London.
5. FREEMAN C & JAHODA M (Eds) (1978) *World Futures: The Great Debate.* Martin Robertson, University of Sussex, UK.
6. GOLDSMITH ERD *et al.* (1972) Blueprint for survival. *The Ecologist* **2**(1).
7. KANE J Intuition, Policy and Mathematical Simulation (private document). University of Columbia, Canada.
8. KANE J (1972) A primer for a new cross impact language. KSIM. *Technol. Forecasting and Soc. Change* **4**, pp. 129–142.
9. LEICESTER C (1972) *Britain in 2001 A.D.* H.M.S.O., London.
10. MARTEL (1977) The growth of growth. *Futures* **9**.
11. MEADOWS DH, MEADOWS DL, RANDERS & BEHRENS (1972) *Limits to Growth.* Universe Books, New York.
12. MESAROVIC M & PESTEL E (1974) *Mankind at the Turning Point.* Dutton/Readers Digest Press, New York.
13. World Bank (January 1972) Report on Limits to Growth.

ADDRESS FOR CORRESPONDENCE: *Professor BHP Rivett, Operational Research, The University of Sussex, Mantell Building, Falmer, Brighton BN1 9RF, UK.*

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 43–53
© Pergamon Press Ltd. 1979. Printed in Great Britain
0306-0483/79/0201-0043$02.00/0

The World Strike Wave

DUDLEY JACKSON
TIM SWEET

The University of Aston Management Centre, Birmingham, UK

(*Received March 1978; in revised form September 1978*)

The years since 1969 have seen a major increase in strike activity in nearly all the countries which collect and publish data on strikes. This wave of strikes has largely been caused by strikes over pay, probably connected both with higher rates of price inflation and a higher incidence of direct taxation on households. In some cases, the changed *level* of strike activity has also led to substantial changes in the *structure* of strikes, as we shall show for the United Kingdom and Australia, two countries which have suffered most from the strike wave.

INTRODUCTION

THE LEVEL of strike activity in the world's industrial market economies increased sharply in the period from, and including, 1969. Table 1 provides some introductory figures on total working days lost in the industrial countries treated as a whole. Because of the vastly greater size of the labour force in the United States, we present figures for the US separately. In the years immediately following the end of the Second World War, the industrial OECD countries (excluding the US) lost on average a total of 31.7 million working days through strikes in any one year. For the period 1951 to 1967 inclusive, this annual average of total working days lost dropped considerably by about one-fifth to 25.3 million. (We are forced to exclude 1968 from this table, because no data is available for France in this year.)

TABLE 1. WORKING DAYS LOST IN WHOLE ECONOMY THROUGH WORK STOPPAGES IN INDUSTRIAL OECD COUNTRIES

	Annual average number in millions of working days lost through work stoppages		
	1946 to 1950	1951 to 1967*	1969 to 1976
Total working days lost			
Industrial OECD countries (exc. USA)†	31.7	25.3	58.8
USA	54.8	29.8‡	41.1

* 1968 is excluded because no data on work stoppages in France is available for that year.

† The full list of countries is: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany (Fed. Rep.), Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Sweden, Switzerland, United Kingdom, United States of America. Note that 1946 to 1950 exclude Austria; and 1946 to 1948 exclude both Italy and Germany.

‡ Including 1968.

Sources: calculated from data in International Labour Office, *Year Book of Labour Statistics:* 1951/52, Table 38, pp. 327–329; 1955, Table 36, pp. 339–402; 1956, Table 35, p. 459; 1958, Table 35, p. 554; 1966, Table 30, pp. 710–715; 1975, Table 27, pp. 798–804; 1977, Table 27, pp. 842–850.

From 1969 to 1976, the total number of working days lost in the 'average' year rose to 58.8 million. This is more than double the annual average for the preceding period.

The figures for the United States are somewhat different. The immediate post-war years were a time of considerable industrial unrest, after which the American total of working days lost fell considerably. Although the US also experienced a 'strike wave' in the period 1969 to 1976, the rise was much less marked and the incidence of strikes remained below that of the immediate post-war years.

The timing of the strike wave may be seen quite clearly in Fig. 1 which charts the year-to-year course of total working days lost in the OECD industrial countries (excluding the United States). The relatively, but not excessively, high level of strike activity in the period 1946 to 1950, may possibly be explained as a consequence of the relaxation of wartime constraints. Then follows a long period of seventeen years of relatively stable industrial relations with the total number of working days lost see-sawing up and down (as do most of the conjunctural variables of that period).

Against this background, the graph's movement in 1969 and subsequent years is most impressive. It is in this period that the annual total of working days lost rises, on average, to 58.8 million.

We may pursue the theme of the rise in working days lost by considering the relationship of working days lost in the 'industrial' sector only of each country relative to the 'industrial' labour force of each country. Dividing each country's working days lost in 'industrial' stoppages of work by the industrial labour force means, of course, that the figures for each country are standardised relative to labour force size and that therefore the resulting statistic is comparable among the various countries.

Table 2 shows the figures for annual working days lost per thousand employees as averaged among all the OECD countries and also as then averaged over the two separate periods 1953 to 1968 and 1969 to 1976. This is an unweighted average, intended to show the average experience of a 'typical' country. The average annual incidence of working days lost doubled between these periods from 367 days per thousand employees to 715 days (including the USA).

Figure 2 charts the year-to-year course of the average number of working days in industry lost per thousand employees and it shows, in even more remarkable form, the strike wave of 1969 to 1976. Again, 1969 appears very clearly as the first year of the strike wave.

So far, we have considered the strike wave simply as an overall phenomenon. Table 2 also shows in some detail how it affected individual countries. Out of the seventeen countries listed only four show a fall in the incidence of strike activity in the industrial sector (France, Japan, Switzerland and Norway). Most of the other countries show fairly substantial rises in the incidence of strike activity: in seven countries

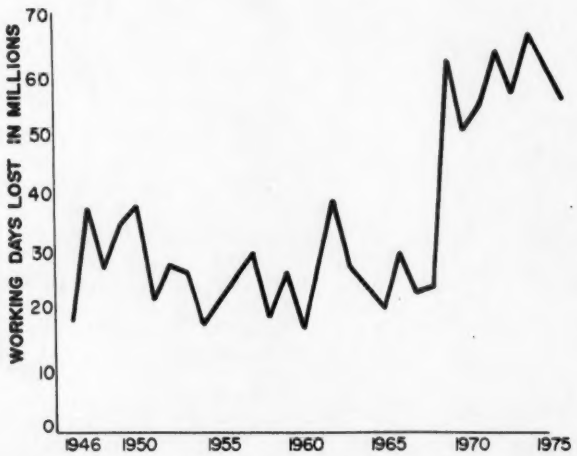


FIG. 1. *Total working days lost through work stoppages in industrial OECD countries (excluding USA).*

TABLE 2. WORKING DAYS LOST IN INDUSTRY PER THOUSAND EMPLOYEES*

Country	Annual average working days lost per 1,000 employees		Index for 1969 to 1976; 1953–68 = 100
	1953 to 1968	1969 to 1976	
Finland	261	1,114	426
United Kingdom	267	911	341
Australia†	409	1,339	327
New Zealand	143	455	319
Canada	780	2,024	259
Italy	893	2,091	234
Denmark‡	349	709	203
Netherlands	39	76	194
Sweden§	32	49	154
Ireland	633	905	143
Germany (Fed. Rep.)	54	66	123
United States‖	1,072	1,309	122
Belgium¶	388	434	112
France**	381	314	82
Japan	348	272	78
Switzerland	8	4	44
Norway	185	81	44
All countries††	367	715	195

* The industries covered are mining, manufacturing, construction and transport.
† Includes electricity and gas; excludes communication.
‡ Manufacturing only.
§ All industries until 1971.
‖ Includes electricity, gas and sanitary services.
¶ Average for second column excludes 1976.
** Average for first column excludes 1968.
†† Unweighted averages.
Sources: calculated from data in: *Ministry of Labour Gazette* October 1963, p. 400; *Employment and Productivity Gazette*, November 1968, p. 917; *Department of Employment Gazette*: December 1974, p. 1110; December 1977, p. 1342.

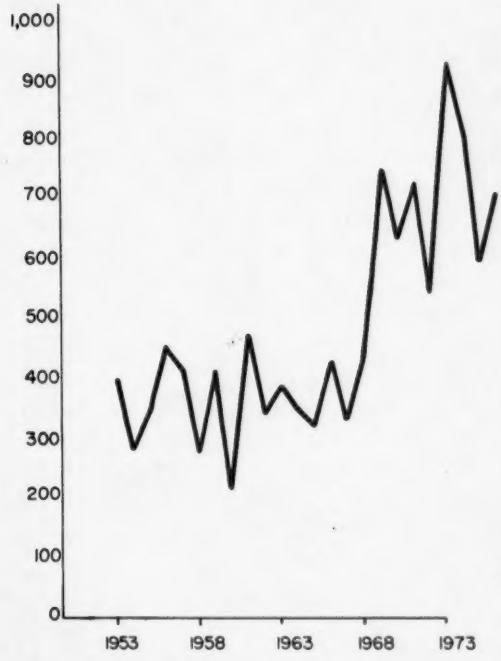


FIG. 2. *Working days lost per 1,000 people employed in industrial OECD countries (including USA).*

the incidence more than doubles. Overall, as we remarked before, the strike wave represents a doubling in the incidence of working days lost.

Table 3 continues the saga of the strike wave, showing how both the number of strikes and also the number of strikers in countries other than the USA about doubled. Again, these changes are quite widespread among countries [7].

From all this, we conclude that the existence of a world strike wave, starting in 1969, is incontrovertible. However, the examination and interpretation of strike data over the long term is a particularly hazardous occupation. The difficulties involved are illustrated by Ross and Hartman's famous thesis of the withering away of the strike, propounded in 1960, but now shown to be inaccurate [5]. Indeed, the strike wave which we have dated may represent a jump to a new level of industrial activity, or may be simply part of a cycle. There have been many econometric and other studies which

TABLE 3. NUMBER OF STOPPAGES IN WHOLE ECONOMY AND WORKERS INVOLVED IN INDUSTRIAL OECD COUNTRIES

	Annual average		
	1946 to 1950	1951 to 1967*	1969 to 1976
Number of stoppages			
All industrial OECD countries (exc. USA)†	6,576	9,609	18,757
USA	4,109	4,054‡	5,459
Number of workers involved in work stoppages (in millions)			
All industrial OECD countries (exc. USA)	6.77	7.25	16.01
USA	2.83	1.95‡	2.50

* 1968 is excluded because no data on work stoppages in France is available for that year.
† Excludes Austria and Germany.
‡ Including 1968.
Source: as for Table 1.

attempt to show a relationship between strike activity and other conjunctural variables, such as unemployment and inflation [1, 2, 6]. The general conclusion which they reach is that strike activity follows a cyclical pattern corresponding to the business cycle. But the magnitude of the strike wave since 1969 is such that it cannot be explained as part of a short or medium term cyclical movement of the economy. The only other period this century when strike activity occurred on a similar magnitude to the last decade and in several countries was in the 1920s. Relative to the whole post-war period, the 1970s strike wave is of considerable magnitude—that is, we are talking not of a minor but of a major change in workers' collective behaviour. This unrest, shown in the labour relations of the world's major economies (or, rather, in those countries where such activity is, more or less, freely permitted and recorded) continued unabated through the years since, and including, 1968. We are, as yet, unable to date the end of the strike wave for it shows no signs of subsiding.

What caused the strike wave? And what effect did it have on the structure of strikes?

STRIKES, INFLATION, AND TAXATION

So far we have assumed a common starting date of 1969 for the beginning of the strike wave. But of course the strike wave did not start in unison in every country. Through statistical procedures and an examination of other evidence, we have established for each country the starting year of its particular strike wave; Table 4 shows the results of this examination [8]. We can see that the era of higher strikes started in Canada, Ireland and New Zealand as early as 1966, and that the strike wave did not catch up with Denmark until 1973 and Japan until 1974. The United States' strike wave, such as it was, started in 1967; and, of course, the French strike wave must be counted as beginning in the famous year of 1968. The strike wave in Italy, the United Kingdom and Australia started in 1969, and spread to the remaining Western European economies in the years following, except for the three countries—Norway, Switzerland, and Austria—which so far have been immune.

During the strike wave all these countries have suffered from higher rates of price inflation and it is natural to think that the strike wave and rapid inflation are causally linked. There are two competing hypotheses: the first is that rising prices, by pressing on workers' standards of living, caused the strike wave; the second is that the effect of strikes was to force up wage rates and thus labour costs per unit of output so producing more rapid price inflation. In either case one is left with the problem of explaining what initiated the process. But of course once the process is initiated, it is obviously a self-reinforcing spiral.

Vol. 7
1979

TABLE 4. ESTIMATES OF DATES ON WHICH COUNTRIES' STRIKE WAVES AND ACCELERATED INFLATION BEGAN

Country	Year in which: Strike wave began	Inflation accelerated
Canada	1966	1972
Ireland	1966	1969
New Zealand	1966	1970
United States	1967	1968
France	1968	1969
Italy	1969	1973
United Kingdom	1969	1970
Australia	1969	1971
Belgium	1970	1972
Netherlands	1970	1969
Germany (Fed. Rep.)	1971	1971
Finland	1971	1973
Sweden	1971	1974
Denmark	1973	1973
Japan	1974	1973
Norway	—	1970
Switzerland	—	1971
Austria	—	1972

Source: T G Sweet, "The connection between strikes and inflation", University of Aston Management Centre Working Paper Series No. 71: available on request.

One way of beginning to examine this very complex problem was to ascertain—again by statistical means and other evidence—the year in which a country's rate of inflation could be said to have begun to accelerate above its previous historical 'norm'. Table 4 shows also our listing of these years.

The remarkable thing is that the year in which price inflation accelerated was nearly always a year *subsequent* to the beginning of the strike wave: in the case of the Netherlands and Japan, price inflation precedes by one year, and in the case of West Germany the years coincide. Otherwise there tends to be a lapse of a year or two between the onset of the strike wave and the beginning of higher rates of price inflation. In the light of this evidence, the argument long since put forward by the OECD secretariat that accelerating wage inflation—i.e. rising unit labour costs—is a major constituent, and perhaps a direct cause, of recent price inflation, achieves a greater significance.

Of course, once the process of wage-price inflation has begun, subsequent rounds of price rises will cause demands for higher wages. Not only this, but the sectional process of collective bargaining implies that any disturbance to established 'fair' wage differentials will be resisted: it takes only one successful strike to set off a chain of demands for comparable pay rises. Wage comparability with similar occupations is much cited in negotiations between employers and unions in all the countries we have examined [8]. In the Australian system of industrial relations the objective of comparative wage justice, by which is meant paying the same wage to people in the same occupation but in different industries, is an important principle of their arbitration system [9]. In this way, one strike producing a wage increase can lead to a succession of disputes in similar occupations.

However, Table 4 still leaves us with the problem of explaining what started the world strike wave in the first place. We have no complete answer to this question and perhaps there is no answer common to all countries. It is often argued that the events in France of May 1968—perhaps like the events of May to July 1789—stirred the aspirations of the common man around the world. Unfortunately for this explanation, strike waves had begun in four countries before 1968, including the United States.

There are many explanations propounded for the inflation process which has affected the

TABLE 5. CHANGES IN THE LEVEL OF PERSONAL TAXATION

Country	Personal income tax and social security contributions as a percentage of GNP		Index 1967–69: 1955–57 = 100
	1955–57	1967–69	
Ireland	4.1	7.3	178
Netherlands	14.2	24.2	170
Canada	6.8	11.2	165
Sweden	15.5	25.3	163
Norway	13.4	21.7	162
United Kingdom	10.1	15.7	155
Italy	10.8	16.1	149
Denmark	11.5	16.6	144
Finland	10.6	15.3	144
Belgium	12.2	17.6	144
United States	12.0	16.5	138
France	14.1	19.1	135
Switzerland	10.4	13.5	130
Austria	14.5	18.5	128
Japan	5.9	7.5	127
Germany	14.8	18.7	126

Source: OECD, *Expenditure Trends in OECD Countries 1960–1980*, Table 19, p. 71.

world's industrial market economies ever since 1945, but with special ferocity in the last decade. The three most familiar explanations are the essentially 'Keynesian' mechanism of excess demand, the 'monetarist' theory of an uncontrolled growth in the money supply, and the idea of increased trade union militancy pushing up wages. These hypotheses have been much tested in the literature on the subject, and we have discussed their relative merits elsewhere [8]. But in terms of the world strike wave all these hypotheses suffer from two distinct disadvantages. First, no single theory can provide an explanation which will fit the experiences of all the sixteen countries, or even of a substantial proportion of them. Second, as we argued above, the magnitude of the upsurge in industrial unrest in the past decade has been so great that theories which previously seemed to explain strike activity are no longer consistent with this sudden and dramatic change in the state of industrial relations.

For the UK, it has been argued that the increasing burden of income taxation and social security contributions played a major role in setting off the wave of strikes in 1969 [3]. Judging by the British government's subsequent use of tax reductions in attempts to mitigate pay demands, the force of this argument has been admitted officially. But could this be a general explanation of the world strike wave?

Table 5 provides some international comparative data which suggests that this may be as near as one can come to a general explanation. Over the period 1955–57 to 1967–69 (just before the general 'beginning' of the strike wave), taxes on, and social security contributions paid by, households, measured relatively to gross national product (and thereby discounting the effects of inflation), increased markedly in all countries for which data are available. In Table 5, countries are listed in descending order of the proportionate increase in the relative incidence of taxation on households. For those countries with overlapping statistics on strikes and taxation, there appears even to be a rough correspondence between the severity of the increase in the incidence of strikes (as measured in Table 2) and the extent of the increased incidence of taxation (as measured in Table 5)*.

This evidence is somewhat impressionistic. We can say little more than that the phenomenon of increased strike activity was preceded by an increasing incidence of taxation on households. The probable connection between the two is, of course, that increases in taxation will affect household *disposable* income, and

* The coefficient of rank correlation is +0.31 for fifteen countries: excluding the case of Norway (whose strike incidence decreased while the tax incidence increased), the coefficient of rank correlation for 14 countries is +0.62.

TABLE 6. STRIKES OVER PAY AND THE WORLD STRIKE WAVE

	Annual average of working days lost through stoppages, in thousands					
	Strikes over pay		All other strikes		All strikes	
	1951 to 1968	1969 to 1974	1951 to 1968	1969 to 1974	1951 to 1968	1969 to 1974
Australia	341	2,241	449	818	790	3,060
United Kingdom*	2,764	10,991	1,059	1,898	3,823	12,889
Italy†	6,200	17,681	3,629	3,508	9,829	21,189
New Zealand	45	105	69	90	114	195
Belgium	289	596	275	158	563	755
Ireland	183	358	101	172	284	530
United States	20,267	34,292	9,534	9,021	29,801	43,313
Switzerland	2.7	2.2	8.2	0.3	10.9	2.5

* The UK data for the earlier period is from 1957 to 1968.

† Italy's reported working hours lost has been converted to working days lost by dividing by the reported length of the working days of 7.78 hours (1969 to 1974).

Source: calculated from official national publications: Australia: *Labour Report*; United Kingdom: *Department of Employment Gazette*; Italy: *Annuario di Statistiche del Lavoro e dell' Emigrazione*; New Zealand: *Prices, Wages and Labour Statistics*; Belgium: *Annuaire Statistique de la Belgique*; Ireland: *Irish Free Trade Journal and Statistical Bulletin*; United States: *Handbook of Labor Statistics*; Switzerland: *Annuaire Statistique de la Suisse*.

this tax effect may assume a more critical significance at a time of rising prices by exacerbating the difficulty of maintaining standards of real consumption.

It is therefore most intriguing to find that the recent increase in strike activity has been largely due to an increase in strikes over pay. Table 6 lists the countries for which data is available on strikes classified by 'cause' of strike—or, rather, by principal reason given for striking. The distinction between strikes over pay and over other issues is not an easy one to draw. The distinction in Table 6 is more properly between 'economic' and 'non-economic' issues. Economic issues cover such things as wages, hours and leave, compared with disciplinary matters or working conditions. There will be some overlap between these two categories, but in the countries cited in Table 6 the distinction is, in our opinion, sufficiently accurate to support the conclusions which we draw. Here we have measured the incidence of a country's strike activity by its annual aggregate of working days lost. In all the eight countries for which such data are available, annual working days lost in strikes over pay have risen remarkably, when the strike wave period (1969 to 1974) is compared with the previous period (1951 to 1968). Australia heads the list with a six-and-a-half-fold increase in annual working days lost through strikes over pay; the United Kingdom is second with a four-fold increase; and Italy's increase is nearly three-fold.

Increases in working days lost through pay strikes in other countries except Switzerland have also been substantial. Among countries which publish data on strikes classified by cause, Switzerland is unique in showing an actual decrease in the incidence of pay strikes during the world strike wave. However, this may be symptomatic of other things in Swiss society, as their incidence of non-pay disputes dropped to negligible proportions.

By contrast with these sharp increases in the incidence of pay strikes, the changes in the average incidence of non-pay disputes were very minor, with the exception of Australia and the United Kingdom where the incidence of non-pay disputes more or less doubled. Indeed, in several countries the incidence of non-pay disputes actually declined during the strike wave.

On average in all the countries listed, working days lost in pay strikes accounted for 53% of annual working days lost in the period 1951 to 1968 (1957 to 1968 in the case of the UK). This average proportion rose to 76% during the strike wave from 1969 to 1974. Such a change in the structure of working days lost is eloquent testimony to the increased importance of strikes over pay.

THE CHANGING STRUCTURE OF STRIKES

For both Australia and the United Kingdom, the countries near or at the top of the list in Tables 2 and 6, the effect of this change in the 'structure' of strikes has been remarkable. Briefly, British and Australian strikes have become much more grand affairs.

Table 7 has been compiled from data in the United Kingdom's Department of Employment *Gazettes* and from the Australian Bureau of Statistics *Labour Reports*. It shows that in the long period of 'normality' from 1946 to 1968, the average British strike lasted only about four working days, involved 402 workers both directly and indirectly in the establishment(s) where the stoppage was taking place, and so caused an average loss of 1372 working days per stoppage. During the strike wave, by contrast, the average British strike lasted much longer: 7.8 working days on average; involving 496 workers, and caused a loss of 4072 working days per stoppage. (Note that this *excludes* the 'political' strikes in protest against the Conservative government's Industrial Relations Act.)

Table 7 also shows that the average Australian strike has similarly become much larger. During the period 1951 to 1968 the average Australian strike lasted 1.8 working days, involved 337 workers (again, directly and indirectly in the establishment(s) where the stoppage occurred), and caused the loss of 608 working days per stoppage. Yet during the strike wave the length of the average stoppage increased slightly in duration to an average 2.2 working days; involved many more workers: 606 on average; and caused the loss of 1321 working days.

There are problems with comparability of statistics between countries, and this is particularly so in the case of strike statistics [10, 11]. Consequently we have confined most of our discussion to the consideration of working days lost, the most reliable statistic for international comparisons of industrial disputes. But the strike statistics of both the United Kingdom and Australia, are sufficiently accurate and consistent to draw some meaningful conclusions from them. The figures show that the average Australian strike tends to be much shorter than the average British strike. This is probably an accurate reflection of the differing structures of industrial relations in each country: in Australia the legal arbitration system puts a premium on short stoppages, whereas in the United Kingdom the system of collective bargaining has resulted in many protracted struggles over pay. This greater length of British strikes, rather than the number of workers per stoppage, explains why British strikes tend to be bigger in terms of working days lost.

Not only has the average strike in the United Kingdom and Australia during the strike wave become longer and, in particular, bigger in terms of working days lost per stoppage, but stoppages have also become more numerous. In the earlier period, the annual number of stoppages averaged 2157 in the

TABLE 7. CHANGES IN 'THE AVERAGE' BRITISH AND AUSTRALIAN STRIKE DURING THE STRIKE WAVE

	Annual average			
	United Kingdom*		Australia†	
	1946 to 1968	1969 to 1976	1951 to 1968	1969 to 1976
Working days lost per worker involved‡	4.2	7.8	1.8	2.2
Number of workers per stoppage	402	496	337	606
Working days lost per stoppage	1,372	4,072	608	1,321

* Stoppages beginning in year; and workers involved and working days lost in these stoppages.
† Stoppages in progress in year; and workers involved and working days lost in year.
‡ This is an approximate measure of the average duration (in terms of working days) of stoppages: it is only an approximate measure because it is not always the case that *all* the workers involved will stop and resume work on the same days.
Sources: United Kingdom, *Department of Employment Gazettes*; Australian Bureau of Statistics, *Labour Reports*.

United Kingdom and 1284 in Australia; during the strike wave the annual numbers averaged 2730 and 2411 respectively. But why were there more stoppages and why did the average stoppage become so much bigger? Part of the answer lies in the changing structure of strikes.

An important similarity between British and Australian stoppages is that the average dispute over pay tends to be much larger than the average non-pay dispute. Table 8 shows that, in the period 1957 to 1968, the average pay dispute in the United Kingdom caused the loss of 2.9 times as many working days than did the average non-pay dispute. Furthermore, during the strike wave, the average British pay strike much more than doubled in size, and caused the loss of 3.8 times as many working days than did the average non-pay dispute. In Australia in the period 1951 to 1968, the average pay dispute caused the loss of 4.3 times as many working days as did the average non-pay dispute. Table 8 shows that during the Australian strike wave both pay and non-pay disputes increased in size.

But Table 8 also shows that the major contribution to the increased *number* of stoppages during the strike wave came from pay disputes. Indeed, in the United Kingdom the average annual number of pay disputes rose by one-third during the strike wave while the average annual number of non-pay disputes actually declined slightly. In Australia the average annual number of pay disputes increased by more than four-fold between the two periods, while the number of non-pay strikes increased by a little more than one-third. It is clear, then, that the total number of stoppages rose primarily because of the rise in the number of pay strikes. Because such strikes tend to be bigger, this had an inevitable effect on 'the average' British and Australian strike.

However, this is not the whole explanation for the change in the size of the average British or Australian strike. Table 8 (and Table 6) show that the effect of the decline in the *number* of non-pay disputes in the United Kingdom was more than made up for by a large increase in the *size* of the average non-pay dispute from 844 working days lost per stoppage to 1569 working days lost per stoppage; this increase in average size meant that the total number of working days lost in non-pay disputes also increased during the strike wave. In the United Kingdom, the increase in working days lost has not been confined to pay disputes. Similarly, for Australia Table 8 shows that while the number of non-pay disputes increased only slightly, the size of these non-pay disputes increased markedly from 417 working days lost per stoppage to 744 working days lost per stoppage. Consequently, as shown in Table 6, the annual average total of working days lost through non-pay disputes in Australia nearly doubled between the two periods. This characteristic of increase in the incidence of non-pay disputes during the world strike wave is shared by the United Kingdom and Australia, but not by any other country in our (admittedly small) sample.

SOME CONCLUDING SPECULATIONS

It is clear that a number of important points

TABLE 8. PAY AND NON-PAY DISPUTES IN THE UNITED KINGDOM AND AUSTRALIA

	Annual average			
	United Kingdom*		Australia*	
	1957 to 1968	1969 to 1976	1951 to 1968	1969 to 1976
Number of stoppages				
Pay disputes	1,143	1,557	220	941
Non-pay disputes	1,268	1,173	1,064	1,470
Working days lost per stoppage				
Pay disputes	2,416	5,923	1,810	2,130
Non-pay disputes	844	1,569	417	744

* See footnotes to Table 7.

Sources: *Department of Employment Gazettes* and Australian Bureau of Statistics, *Labour Reports*.

arise from all these findings. As yet, the strike wave shows no clear signs of subsiding: the metaphor, "a strike plateau" may become a more apt turn of phrase. It is clear that management may have to reckon in future with a greater readiness to strike on the part of workers, particularly in support of pay claims. The statistics on the changed structure of strikes in the United Kingdom and Australia show also that, in these countries at least, the average strike may now be much larger (in terms of working days lost per stoppage) than it previously was; i.e. the economic cost to employers of strikes has increased considerably.

Given the overwhelming predominance of pay strikes, it is natural for attention to concentrate on these. In view of the troubled inflationary situation of these years, coming on top of a trend towards increased taxation of household income, it is hardly surprising that the incidence of strikes over pay increased so markedly.

The remedy for this situation is not in the hands of any single employer: the onus is far more upon the government to ensure that the burdens laid upon household real disposable income by inflation and taxation are as light as possible. The current use by the British government of income tax 'reductions' to compensate for inflation are policies in exactly the right direction. The use of national 'prices and incomes policies', of whatever form, are also necessary policies, and these are, in our opinion, likely to be a permanent part of the 'scenario' of economic policy in the future.

The worsening severity of non-pay disputes in the United Kingdom and Australia is less easily explained and is, perhaps, the most disturbing feature of these countries' strike waves. It is not likely that national policies could do much about this sort of strike activity: here the onus seems to rest upon the individual employer.

Perhaps the increasing militancy shown in pay disputes during the strike wave has somehow or other 'spilled over' into the conduct of non-pay disputes through a sort of 'demonstration' effect. In view of the fact that only two countries among eight have shown this feature, such an explanation would not appear very convincing.

Non-pay strikes cover such matters as: working conditions and facilities; demarcation disputes; redundancy, dismissal and disciplinary issues; levels of manning and work allocation; recognition of, and facilities for, trade unions. It is not surprising that strikes over such essentially 'sectional' or small group, matters should, on average, be so much smaller than disputes over pay, which are likely to involve a large proportion of a company's employees in most of its plants. But questions must now be raised as to the adequacy of the facilities, whatever they may be, in the United Kingdom and in Australia for settling such disputes.

It is a frequently-heard criticism of the Australian system of dispute arbitration that it cannot cope at the local level with such matters. The arbitration system in Australia is geared mainly to dealing with disputes relating to pay and hours and to dealing with disputes at a State or national level.

In the United Kingdom, the past decade or so has seen substantial improvements in various institutional arrangements for dealing with disputes over redundancies and dismissals. The legislation establishing the industrial tribunals, and subsequently extending their jurisdiction, has been widely used [4]. Additionally, the government promulgated a code of industrial relations practice, designed to lay down at least minimum standards. However, the 'returns' to this legislation and other official prompting seem to have been disappointing, to say the least. It is not easy to make long run comparisons because of a change in the classification by 'cause' of strike (or, rather, reason given for striking) introduced in 1973 by the Department of Employment. Nevertheless, the indications are that the years after 1969 saw a large increase in the United Kingdom's working days lost through disputes which should, in theory, have been handled by the new 'judicial' machinery [12]. It seems to us that there can be no real substitute for managerial initiative to reduce these sorts of labour relations disputes.

Acknowledgement—The authors gratefully acknowledge Social Science Research Council financial support for this research.

REFERENCES

1. Bentley P & Hughes B (1970) Cyclical influences on strike activity: the Australian record 1952–1968. *Aust. Econ. Papers*, **IX**, 149–170.

2. GOODMAN JFB (1967) Strikes in the United Kingdom: recent statistics and trends. *Int. Labour Rev.* **95**(5), 465–481.
3. JACKSON D, TURNER HA & WILKINSON F (1972) *Do Trade Unions Cause Inflation?* Cambridge University Press.
4. JACKSON D (1975) *Unfair Dismissal. How and Why the Law Works*. Cambridge University Press, Chap. 4. pp. 46–72, and Tabs. 6, 7, 8 and 12 on pp. 51, 52, 59 and 63.
5. ROSS AM & HARTMAN PT (1960) *Changing Patterns of Industrial Conflict*. John Wiley, London.
6. SAPSFORD D (1975) A time series analysis of UK industrial disputes. *Ind. Rel.* **14**, 242–249.
7. SWEET TG & JACKSON D (1977) *The World Strike Wave 1969 to 197?*. Working Paper Series No. 63. University of Aston Management Centre, Birmingham, UK.
8. SWEET TG (1977) *The Connection between Strikes and Inflation*. Working Paper Series No. 71. University of Aston Management Centre, Birmingham, UK.
9. SWEET TG (1978) *Australian Strikes and Arbitration: 1822 to 1976*. Working Paper Series No. 87. University of Aston Management Centre, Birmingham, UK.
10. SWEET TG & JACKSON D (1978) *The Classification and Interpretation of Strike Statistics: an International Comparative Analysis*. Working Paper Series No. 97. University of Aston Management Centre, Birmingham, UK.
11. SWEET TG (1978) *Australian Strike Statistics 1913 to 1976*. Working Paper Series No. 89. University of Aston Management Centre, Birmingham, UK.
12. SWEET TG AND JACKSON D (1977) *The Changing Nature of British Strikes*. Tabs. 9 and 10 in Working Paper Series No. 79. University of Aston Management Centre, Birmingham, UK.

ADDRESS FOR CORRESPONDENCE: *Prof. Dudley Jackson and Tim Sweet Esq. The University of Aston Management Centre, Maple House, 158 Corporation Street, Birmingham B4 6TE, UK.*

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 55–60
© Pergamon Press Ltd. 1979. Printed in Great Britain
0306-0483/79/0201-0055$02.00/0

Modelling the Use of the Manager's Time

DANIEL J BROWN

The University of British Columbia, Canada

This paper starts from two problems: the lack of use of queueing theory (and OR models in general) and the need for greater personal time effectiveness expressed by managers. The application of waiting line theory to the general managerial role is proposed and the most salient assumptions in such an application are explored. A simple example is undertaken which results in tentatively promising evidence. Further questions and possibilities are raised.

HOW TO MODEL A MANAGER

A RECENT survey of applications of management science techniques by Ledbetter and Cox [8] reveals the relative paucity of some familiar models, notably queueing theory. Byrd [4] raises the appropriateness potential of queueing theory specifically and Bhat [2] responds by saying that successful application depends upon a knowledge of a problem area. At a time when a general debate continues as to the usefulness of OR/MS models, it would appear that a new application of queueing theory would be most welcome.

Concurrently, corporate managers have continually expressed their need to manage their personal time more effectively, to reduce time lost, and to allow time for important work to be done. This is especially true of managers who come in contact with many people per day, as indicated by Mintzberg [10] who stresses the chaotic aspects of managerial life. This is a problem of great personal meaning which most managers will acknowledge readily.

The purpose of this paper is to suggest a marriage between queueing theory and managerial life and to demonstrate, at least initially, how a courtship might be undertaken. If it is possible to vary key factors in managers' daily existence which influence the amount of time lost, either on the part of individual managers or the persons with whom they come in contact, then perhaps it is possible to reduce costs or increase performance. The *modus operandi* is simply to consider the administrative process as a set of waiting lines of people directed toward the manager who is the bottleneck through which they must pass in receiving some kind of service. Most administrative roles conform to this description to a considerable degree. An executive services several queues: face-to-face short service, appointments, telephone calls, and correspondence, each of which may contain lower managers, customers, superiors, clerical workers, and fellows of the same rank. For each waiting line, there is a priority system to determine who is to be served in what order. For instance, paperwork is usually shuffled according to some kind of priority before making responses. The lines also have their characteristic arrival and service rates.

Once viewed in this perspective, the process may be observed, appropriate measures taken, and inputs may be varied to see if time may be saved or costs reduced. One output is the amount of time that the manager will be 'idle' or free from service duties. This will vary during the day and has implications for workloads with other queues (a lack of people and telephone calls may permit progress to be made on paperwork). Another output is the total time all persons will have to wait in their respective lineups. The associated costs may be balanced so that the overall costs to the corporation may be minimized. A further result is the qualitative benefit of a greater understanding of the particular managerial role to which the model is applied.

However, the thrust of this paper is an exploration of the appropriateness of various queueing models for the management role *in general*. The challenge, as with all practical applications of mathematical models, is to posit a general model which makes as few simplified assumptions as possible on the one hand and provides a reasonable representation of reality

on the other. In the trade-off between parsimony and reality, it is not expected that the reader will always agree with the author as to which assumptions may be waived and which others are seen as a crucial part of any general model of the role. It is convenient, although arbitrary, to divide the conceptualization of the queueing problem into assumptions dealing with arrivals, queue discipline, and service.

ARRIVALS

A start may be made by entertaining a simplifying assumption which deals with the concept of group arrivals. While standard queueing theory addresses the behavior of *individuals* as they wait in line, it is just as easy to consider *groups* as the units of analysis. For instance, workers or visitors may come in pairs or triads to see the manager with regard to a single issue. If group size is permitted to vary from one, two, three and upwards, then the description of the process fits reality much better. The idea of group arrivals has been successfully applied to the airline reservation problem where passengers frequently arrive in groups of family or friends numbering from one to seven or more [9].

A question of crucial importance to any application of queueing theory relates to the problem of equilibrium. Here, there are basically two choices. The first is the presumption that the queueing process has been operating for a long enough time to be independent of the starting effects which are based on the fact that no one is usually in the waiting line when the day begins. The payoff of this particular assumption is a considerable reduction in problem complexity. The cardinal danger which appears to have been ignored by a number of investigators (as indicated by Baht, [3]) is that equilibrium may not be achieved quickly but may even require a number of hours. It appears quite plausible that the opposite assumption of non-equilibrium would be the accurate one for most administrative situations. Another reality is the variation in the average arrival rate of clients across the day and also across the week. Virtually every manager is accompanied by periods of high demand and low demand for service. The usual way to deal with this problem is to simply posit that the average arrival rate is constant for a specific interval (usually labeled the busy period). Providing the manager's day may be divided into intervals of reasonable lengths of different mean arrival rates, there appears to be little problem.

The most commonly-encountered distribution of arrival times is the exponential, the one invariably discussed in textbooks which introduce queueing theory. But as Baht [3] warns, the exponential assumption is frequently erroneous and should be replaced for one which is more general. For a discussion of the generalities of the arrival time problem see Cox and Smith [5].

Appointment systems are a favourite method for controlling the arrivals of persons for service. Such systems are subject to manipulation by the manager and his or her secretary. Bailey [1] reports the results of appointment experimentation in the case of long queues awaiting medical service from physicians. A similar data-based analysis is presented by Jackson, Welch, and Fry [7]. Although appointment books reduce some complexities for the executive, they permit persons to wait without others being served. This is the case when a subordinate makes an appointment to suit his or her own convenience when the manager may be free at some time between when the appointment is made and when it is met. An initial attack on the difficulty may be made by considering different time ranges of queues. Persons who make appointments ranging over several days can be considered as not waiting and different from persons who simply request short service in an immediate time frame without the aid of the appointment system. Some persons will elect not to joint the queue, probably because the manager is busy. The same outcome would take place if the secretary restricts the length of the queue to a certain maximum number and suggests to newcomers who would further add to the waiting line that they return at a later time. Both these possibilities occur in the queueing literature, and are quite plausible in a managerial setting. Since the four processes (short service, appointments, telephone calls, and correspondence) represent major divisions of the work, it may be necessary to consider them as separate queues, to deal with them in terms of rules of priorities, or both.

Vol. 7
1979

QUEUE DISCIPLINE

It is in this section that the conceptualization of the manager's role has probably the greatest impact on the specification of a queueing model. As alluded to above, one of the more immediate questions which may be raised relates to the division of the role into four aspects of short service, appointments, phone calls, and correspondence which tends to suggest that there are four separate lineups. However, such a structure presents the problem not only of determining priority within a given line, but also priorities across lineups. For instance, are phone calls always considered a higher priority than the personal consultations? And if they are, what is the priority system which decides between the next person waiting and the phone call which arrives and has accumulated no waiting time? Another priority issue is the problem of emergencies. At any particular time, the manager may be faced with a situation which demands his or her immediate attention regardless of present circumstances. Riots or wildcat strikes (other forms of group service) are neither scheduled by appointment nor do they take their place in line. Under such conditions (hopefully not too frequent), considerations about the service of queues of others are normally suspended.

The problem of non-emergency priorities may be assumed to be decided upon a first-come-first-served basis. Once this assumption is made, it is not difficult to vary it to incorporate slightly more complex priorities such as pre-emptive. For instance, while the secretary is being given short service on clerical matters, a superior may successfully request immediate short service. Another priority system is that involving preference but which is non-pre-emptive. In this situation, favoured persons would be permitted to move to the head of the line and are serviced as soon as the previous client is finished. The presence of favoured people raises a natural question as to how many kinds of persons are discernible. And just what is meant by 'kinds of people'? It may be that the only feasible way to answer this question is to use a criterion of importance based upon the particular business which clients wish to transact with the manager. The comparatively arbitrary classification of persons into *a priori* groups may be avoided in this way.

Another issue is the avenue which persons have used in order to gain access to the manager. It appears to be a matter of some agreement among managers that paperwork is one facet of their job which is often neglected and at times ignored by design. This and common sense would suggest that the time frame of reference for paperwork is much more extended than that pertaining to consultations and telephone conversations. In short, letters requiring answers may wait a week, be shuffled according to their own priority system, and be serviced at a relatively slow rate.

Two other relatively minor problems persist. One is the possibility of queue jumping, where individuals or groups may switch from one lineup to another (assuming more than one line). This could be the case where the telephone and the short service lineup are seen as two separate queues and when a person or group perceives a shorter waiting time through the telephone rather than the short service lineup. The other problem relates to reneging, the abandonment of a queue once a person has taken a place in line. These two possibilities may be incorporated if appropriate.

SERVICE

The concerns which relate to the manner in which service is offered are intimately connected with the personal style of the manager. For instance, although a common assumption of most queueing models is that arrival and service rates are independent, should the manager be in the habit of accelerating his or her average rate of service because of an increase in the average rate of arrival, the assumption is violated as suggested by Lee [5]. Another manager-specific feature of service is the average service time, which doubtless varies considerably. This time may undergo some shift during the day as a result of fatigue or personal habit. If such variation is large, then it could be taken into account in the same manner as were the variations in mean arrival rate: by allocation of parts of the day or week into 'fast' and 'slow' periods.

As with arrivals, a standard consideration is the shape of the service time distribution, discussed by Cox and Smith [7]. It may be beneficial to invoke a cyclic assumption, cyclic service being defined as a situation wherein

persons may be served more than once. Such would be the case where persons who work closely with the manager gain his or her audience several times a day. This kind of service is most apparent in the office when subordinates, both in groups and singly, are served again and again. As an initial assumption, however, cyclic service is quite complex and would be applicable only if the periodicity of service appears to occupy an important segment of the administrative process.

A VERY SIMPLE EXAMPLE

The above model of the general administrative role indicates that a queueing model is to some degree appropriate but that the application is a relatively complex one. This discussion has favoured a model which involves appointment systems, may not be in equilibrium, has four waiting lines with priorities, and permits variation in arrival and service rates across the day and week. The complexity of the model precluded the use of an analytic method since the particular model's equations could not be located or developed within a reasonable time. However, once this kind of application acquires a standard form, such as the frequently-described multiple-server queue, there should be no difficulty in producing results which are free of the usual worries of simulation methods.

So that the model outlined above might be properly demonstrated to be applicable, it was decided to model a chairman of a university department during his office hours for a typical week. His clientele included a range of people: students, secretaries, professors, and administrators. Since the amount of mail was extremely small, three queues of telephone, short service, and appointment were used. Priorities were pre-emptive, in the above order, with succeeding telephone calls pre-empting earlier ones. Otherwise, the first-come-first-served rule applied. The time between when an appointment was made and when it was begun was ignored, since it was not considered lost time. Waiting time for clients and idle time for the chairman was considered lost in this case since highly productive work could not be carried out in the busy office setting.

Service was monitored during the 12 hours for which the office was open. The task was split between the secretary, who noted the time at which each person arrived at the office, and the chairman, who recorded the starting time, stopping time, and queue of each person served. For the chairman, this was very similar to a simple list of his activities and their associated times.

The analysis proceeded as follows: first, the input distributions were determined. These were the arrival and service distributions for each of the telephone, short service, and appointment queues. Then the chairman was simulated using a table of random numbers to generate the client arrivals on all three queues concurrently. Next, the following information was determined for each client: arrival time, service starting time, service ending time, waiting time before and during service, and chairman's idle time. Total waiting time for all clients and total idle time for the chairman were then calculated. Since the major factor over which the chairman had control was the length of his office hours, the analysis was simulated for different numbers of hours, ranging from 60% up to 130% of the original 12.

Cost determination was carried out in the following way: the total waiting time lost by the clients for each analysis was multiplied by $10, the estimated value of an average client's hour in salary and benefits, to arrive at the cost to the client. The total idle time lost by the server for each analysis was multiplied by $20, the estimated value of a chairman's hour in salary and benefits. To this figure was added the number of hours gained or lost (also multiplied by $20) so as to allow for the extra hours spent out of the office or the additional hours spent in the office when the number of office hours was varied. The total of these two costs then gave the chairman's cost. The overall costs were then found by summing the total client and total administrator costs for each set of office hours.

Results are portrayed in Table 1, where the overall cost is shown as a result of various office hours. Note the apparently U-shaped function. It was then possible to determine a more efficient number of office hours, which were less than those presently being incurred. The savings of a move to 80% of office hours was found by the difference between the two costs, or $56 per week for this example.

If these savings are permitted to accumulate

Vol. 7
1979

TABLE 1. TOTAL COSTS RESULTING FROM VARIOUS OFFICE HOURS

Office hours open	Percent of present time	Total costs $ per week
7.2	60	101
8.4	70	81
9.6	80	67
10.8	90	96
12.0	100 (present)	123
13.2	110	180
14.4	120	213
15.6	130	248

over the course of 30 weeks, a normal university year, they grow to $1680. If the administrator is not a department chairperson, but in fact holds office hours for most of the week perhaps 36 hours, then the savings become $5040. If the university is one in which there are 50 chairpersons and 50 other administrators, then the total saving becomes $336,000 per year. This speculative result is extrapolated from just one example and is, of course, suspect. All it does is to raise the possibility of savings through increased efficiency. Note that this kind of saving is not reflected in the budget. Rather, it represents an increase in productivity through the acquisition of lost time. Presumably, it should result in a greater performance for the salary, where the salary input is already allocated.

DISCUSSION

The above analysis could be extended still further. If it is demonstrated that the application of queueing models is fruitful in terms of understanding and enhancing the efficiency of individual administrative roles, then it is quite possible that it would be worthwhile to consider the extension of these models to clusters of positions. Systems would involve interactions among roles such as that which is demonstrated by grievance processes which involve successions of servers through which an individual client must progress. The concept of an administrative system with phases also appears to be useful, wherein short service encounters, appointments, telephone conversations, and correspondence are each important. Servers at one time become customers at another, as shown by the sequence of meetings in which many executives find themselves involved.

One set of important questions which the foregoing application uncovers concern values. Eilon [6] asks "Is there a conflict between the economic performance of an enterprise and the quality of working life?" This issue arises a number of times. For instance, is it desirable for an administrator to function in a queueing situation which permits no 'idle' time? The quality of service has not been considered, although it is within the control of the manager to shorten service times and hence reduce time lost by those waiting. Occasionally, the act of keeping a certain client waiting may be a sign of administrative acumen, rather than inefficiency. Such problems as these require acknowledgement and understanding.

The application of waiting line theory to the actual process of management has two additional benefits beyond demonstrating the usefulness of models to both operations researchers and management. One payoff is that such an application, because of its personal relevance, may have considerable potential as a teaching device. It is difficult to contemplate the application of a linear programming model involving hundreds of variables to the oil industry. It is easy to understand the relevance of arrivals, priorities, and service in our lives regardless of lack of work experience. Based on daily matters, these concepts can bridge the gap between teacher and student. They have empirical referents; they come to have meaning.

The other outcome of the fusion of queueing and administration is that it represents a considerable departure from the usual association of operations research and large-scale problems. Naturally, the bigger the problem, the more sizeable are the percentage and absolute savings. But the world of the small is not without its beauty or necessity and according to Schumacher [11], part of the future which we will all inherit. Perhaps the 'OR of the small' is worthy of some consideration.

Acknowledgements—The author would like to acknowledge the helpful contributions of R Oliver Gibson, Herbert Furash, Marilou Healey, Ian Housego, and Charles Mlynarczyk.

REFERENCES

1. BAILEY TJ (1964) *The Elements of Stochastic Processes with Application to the Natural Sciences*. John Wiley, New York, pp. 147–149.
2. BHAT UN (1978) The value of queueing theory—a rejoinder. *Interfaces* **8**(3), 27–28.
3. BHAT UN (1969) Sixty years of queueing theory. *Mgmt Sci.* **15**(2), B280–294.
4. BYRD J JR (1978) The value of queueing theory. *Interfaces* **8**(3), 22–26.
5. COX DR & SMITH WL (1961) *Queues*. Methuen, London.
6. EILON S (1976) The quality of working life (Editorial). *Omega* **4**(4), 367–373.
7. JACKSON RRP, WELCH JD & FRY J (1964) Appointments systems in hospitals and general practice. *Opl Res. Q.* **15**(1), 219–237.
8. LEDBETTER WN & COX JF (1977) Are OR techniques being used? *J. Ind. Engng* **9**(2), 19–21.
9. LEE AM (1966) *Applied Queueing Theory*. St Martin's Press, New York.
10. MINTZBERG H (1974) *The Nature of Managerial Work*. Harper & Row, New York.
11. SCHUMACHER EF (1973) *Small is Beautiful*. Harper & Row, New York.

ADDRESS FOR CORRESPONDENCE: *Daniel J Brown Esq, Centre for the Study of Administration in Education, Faculty of Education, The University of British Columbia. Vancouver, B.C. V6T 1W5, Canada.*

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 61–66
© Pergamon Press Ltd. 1979. Printed in Great Britain

0306-0483/79/0201-0061$02.00/0

An Application of Catastrophe Theory to Management Science Process

RJ GRAHAM

J SELTZER

La Salle College, Philadelphia, USA

(*Received September 1978; in revised form October 1978*)

Practicing management scientists often complain of the alleged irrational behavior on the part of the managers they are trying to serve, particularly when a manager suddenly and unexpectedly shifts behavior from one of support to one of resistance. The authors feel such behavior is only believed to be irrational because the management scientist is using the wrong mental model when projecting past behavioral patterns into the future. This paper attempts to solve this problem by using the newly developed catastrophe theory to develop a different model where sudden shifts in behavior are considered perfectly rational and explainable. The basic implication that is drawn from this new model is that successful implementation of management science depends on a sequence of interactions with the ultimate model user and that care in the structuring of these interactions can greatly enhance the probability of eventual user acceptance.

INTRODUCTION

MANY PEOPLE who have been involved with the implementation of management science techniques have expressed a frustration resulting from sudden and unexpected reversal of attitudes exhibited by some participants in the management science process. This is, the people collaborating with the management scientist may be in seeming harmony when there is a sudden 'falling out' of one or more participants and behavior turns from cooperative to resistant for no apparent reason. This change in behavior often impedes the eventual use of the model being developed and may cause a technically excellent model to be abandoned for reasons that are seemingly beyond the control of the management scientist. The purpose of this article is to examine the possibility of using the recently developed 'catastrophe theory' as a framework for modelling and understanding such rapid changes in behavior. To achieve this purpose we will begin by describing some of the relevant characteristics of both management science process and catastrophe theory. The theory will then be used as a framework for describing some of the discontinuous behavior often found in the course of implementing management science models. Several implications for the practice and study of the management science process can then be drawn from the catastrophe theory framework.

MANAGEMENT SCIENCE PROCESS

With the available history of successes and failures in the application of management science (MS) techniques, many practitioners and researchers have attempted to delineate the factors that helped or hindered success in implementation, as well as to outline a process by which others could learn from their mistakes [1, 4, 7, 10]. Recently, the term 'management science process' has been used to describe such a sequence of steps that may help to increase the probability of successful implementation of the MS model [2, 9]. The suggested processes vary from author but most contain the following as vital steps.

1. Problem identification—an intricate process by which an irritation or feeling that all is not well is translated into a problem statement (or statements). This step should include identification of all who will be affected by the solution to the problem (3).

2. Model formulation—being the actual formulation of a model which can be used as an aid in solving/better defining a problem listed in step 1.

3. Model solution—the process of deriving a set of solutions or suggestions from the model formulated.

4. Reality testing—using the results from step 3 on a test basis to ascertain how well the model fits reality.

5. Sustained usage—the continued and regular use of either a particular model or the MS function in an organization.

An important concept to be gleaned from the process shown above is that construction of a model is only one part, and not necessarily the most important part, of successful implementation of MS [1, 4, 10]. Another concept is that implementation does not begin after the model is completed but rather is the aim of the entire process [4]. The successful management scientist should be aware that every step in the process affects the probability of implementation so that concentration on one step to the exclusion of others may seriously decrease the probability of acceptance.

Following all of the steps outlined above, however, does not insure a successful experience. Many structural and behavioral factors such as top management support [6], cognitive style of the manager [8], style of the MS group leader and place of the MS group in the organization [7] and the like have been suggested as additional factors influencing the success of MS endeavors. Rather than speculate on which of these factors may be more important and under what circumstances, we have chosen to take a different view of the behavioral aspect of MS process.

Any process in a social organization will consist of a series of interactions with people. From the steps outlined above it should be clear that the management scientist will have different levels of intensity of interaction with various people during different phases of the process. The interest and involvement of various people with a stake in the solution to the problem being addressed tends to 'heat up' and 'cool down' as the process proceeds. The usual mental model of such a sequence of interactions is that they are continuous and cumulative such that a series of favorable interactions would predict that subsequent interactions would also be favorable. Any sudden shift in allegiance belies this continuous notion and represents a discontinuity which has been termed a 'catastrophe'. We now turn to an explanation of catastrophe theory as an aid in modelling such a discontinuous process.

CATASTROPHE THEORY

Vol. 7
1979

The recent development of catastrophe theory in the field of mathematics provides a method for describing a number of events that could not be analyzed with previous theories. For example, differential calculus is useful for processes that are continuous and involve the same magnitude of marginal change in the dependent variable regardless of the sign of the marginal change in the independent variable. As such, differential calculus cannot be used to analyze discontinuous processes, particularly when they involve different behaviors of the dependent variable depending upon the direction of change. The speed of a falling object is affected by the continuous pull of gravity and the fall of a body can be easily modelled using differential calculus. But consider other phenomena in which sudden shift or jump or change occurs. For example, a rubber band stretches until breaking, a water pipe bursts, a machine operates and then jams. Catastrophe theory has recently been developed to analyze such discontinuous events, and a full description is given by Zeeman [11].

A '*catastrophe*' can be defined as a sudden disruption that occurs in a continuous process, where a '*behavioral surface*' of possible actions can be identified. Variables that have an impact on the person's position on the behavioral surface are called '*control factors*'. For example, a person begins to laugh and suddenly the laughter turns to tears. The change could be called a catastrophe and the laughing and cry-

ing considered as points on a behavioral surface. Other points might include: hysterical laughing, chuckling, sadness, sobbing and a neutral position of no strong emotion. We can hypothesize that the control factors are feelings of grief and willingness to be expressive of the emotions.

There are five essential elements of a *catastrophe* that need to be mentioned:

1. *Bimodality*—relating to two behaviors that can occur for a given point on the control surface.

2. *Sudden transition*—signified by a disruption, as a person shifts from one portion of the behavioral surface to another.

3. *Hysteresis*—signifying that a transition from one state to another does not imply that the reverse can occur.

4. *Inaccessible region*—refers to a portion of the behavioral surface that cannot be reached from other portions of the surface.

5. *Divergence*—where lines on the surface diverge and the resultant behaviors become quite distinct (near points N and O).

APPLICATION TO MANAGEMENT SCIENCE PROCESSES

In order to use the model to describe discontinuous behavior, we first define the control factors, which are (1) level of involvement in the process and (2) perceived benefit or threat due to the MS (Management Science) project. Much has been written concerning the need for total involvement of key personnel and there is an unstated assumption that this involvement will increase a person's perceived benefit and thus be tantamount to success. Perceived benefit or threat, however, originates from a variety of sources of which involvement may be the least important. The person may perceive the MS model as altering the information flow and power structure in an organization and thus will behave according to whether he feels he will be better off or worse off after implementation (i.e. have more or less power). Intensive involvement may only intensify any negative feelings towards MS implementation. We thus feel that involvement and perceived benefit are like job satisfaction and productivity; they may be correlated, but one does not necessarily cause the other. For this model we

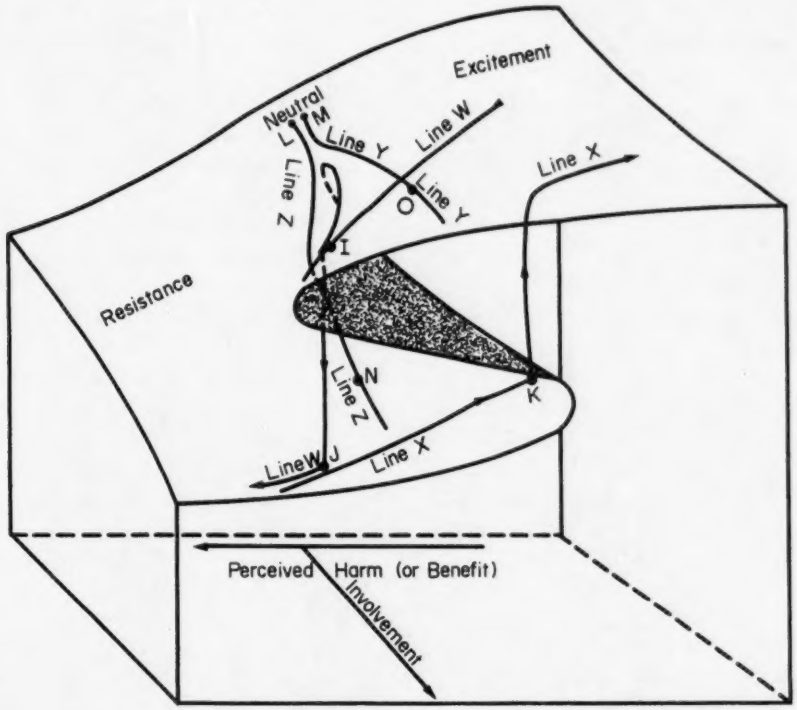


FIG. 1. *Application for Management Science Process (Adapted from Zeeman [11]).*

are thus assuming the factors to be independent.

The catastrophe model of behavior is shown in Fig. 1. The behavioral surface describes the response to a change for an individual. Three points will be identified: resistance, neutral and excitement, although many other points along the surface could also be described. An individual may attempt to interfere with, block or hold back a change or the same individual may support, encourage and push for a change. An individual's sense of the expected benefit or harm that will occur as a result of the change will have a direct effect on the location on the behavioral surface. It is important to note that this is the perceived impact as opposed to the actual impact, since the behavioral response will be based on information that the individual has available. The implication is that as the person gains new information about the change or possibly about the attitudes of others in the organization concerning the change, the behavioral response can also change (including the possibility of a catastrophe). The second control factor is involvement with the management science process. As shown in Fig. 1, this factor is perpendicular to the first control factor and runs directly through the cusp.

Taking the five features of the cusp catastrophes, the model can be explored further. In a situation of high involvement, a small change in perceived threat or benefit can cause a large change in behavior (a sudden transition, i.e. catastrophe). This is observed in an organization as a supporter of a project suddenly becomes resistant (following in Fig. 1 along line *W* from point *I* to point *J*). Further, in this situation, a small change in the direction of perceived benefit is usually not sufficient to convert the detractor back into a supporter. This concept of hysteresis is shown in that on Fig. 1, along line *X*, the transition is not near point J, but rather at point K. The inaccessible region of the behavioral surface occurs in that as people become more involved in the implementation process, they find it harder to maintain a neutral position, but rather become positive or negative. With high involvement, there is a large inaccessible region and only excitement or resistance remains possible. Finally, the idea of divergence can be indicated with lines *Y* and *Z* on Fig. 1. With low involvement, it is reasonably easy to move an individual from point *L* to point *M*. As the person becomes more involved with the implementation, a much greater effort is required to move from point *N* to point O.

IMPLICATIONS

Based on the discussion of management science process and the formulation of the involvement dynamics as a catastrophe model, the following implications can be drawn for both the manager and the management scientist.

1. Formulation of the management science process as a sequence of intensities of interactions is probably more universally applicable than any particular process developed as distinct stages. Practitioners can rarely specify which stage they are in and process is seldom linear and cumulative. Developments in one stage will often force a reversal to a previous stage when the stages are contrasted to reality.

2. The most important portion of any involvement is at the beginning of that involvement. The difference of perceived benefit is very small between point *L* and point *H* as shown on Fig. 1. Assuming the same intensity of involvement, however, the difference in behavior between point *N* and point *O* is quite significant. This formulation supports the notion that "first impressions are lasting".

3. A person with fluctuating levels of involvement in a project cannot be assumed to exhibit the same support (or resistance) with each reintensification of involvement. A highly involved person at point *O* may revert to point *M* with a decrease in involvement. A barely noticeable change in perception of benefits may move the person to point *L* during times of low involvement. With reintensification of involvement, the same person may move to point *N* and be resistant. This then is an extension of implication (2) above in that care must be taken both in the initial involvement and at the start of any reinvolvement.

4. Once there is a change in attitude from assistance to resistance (i.e. catastrophe), due to a negative change in perceived benefits, it

Vol. 7
1979

takes a much larger positive change in perceived benefits in order to reinstate the original excitement. As an example, consider a person at point *I* on Fig. 1. With a slight change in perceived benefits, that person could experience a catastrophe and move to point *J*. The management scientist may realize what caused the change in perception and move the perception back to its original point on the control surface. This movement, however, will not be enough to cause a catastrophe in a favorable direction (i.e. reinstate excitement). The perceived benefits must be moved all the way to point *K* in order to cause the desired catastrophe. Thus, catastrophe theory supports the idea that "an ounce of prevention is worth a pound of cure". It also supports the biological fact that once you are cured of a malady you are much healthier. The person recovered from point *K* will be a stronger supporter than a person continually hovering about point *I*.

5. More is not always better. The conventional wisdom of implementation articles is to have as many people as possible involved in the MS effort from the beginning. This is fine as long as all are supportive of the effort. Those who are resistant, however, may have negative effect on the project and it may be difficult to move them to positive behavior while involvement is high. The theory suggests that in this situation, it may be better to allow involvement to decrease (move from point *N* to point *L*), work on changing the perceived benefits while involvement is low (move to point *M*) and then reintensify the involvement.

6. The ability to maintain a neutral attitude towards the project decreases with increased involvement. This is the application of the inaccessible region. The person must move one way or the other and neutrality becomes impossible.

7. Two people with identical perception of benefits and level of involvement may exhibit entirely different behavior. Another piece of conventional wisdom from the implementation literature is that a cost/benefit analysis of a project will win support for project. The theory suggests that for some people this may not be enough. If a person was very negative at the onset, a revelation of benefits may only move them to point *N*. Others with the same perception (point *O*) may exhibit excitement over the project but more is needed for the person of point *N*. Thus it is important to know not only where a person is on the control surface but also where they came from.

CONCLUSION

We have argued that the management science process is best viewed as a sequence of interactions. Success in implementing the MS model was viewed as mostly a function of successfully managing the interfaces to promote cooperative behavior. The control variables were assumed to be level of involvement and perceived benefit and it was suggested the management scientist should concentrate on these variables to increase the probability of success. Most of the paper considered the interface between the management scientist and one other person. This was for illustrative purposes only as the model will accommodate as many people as necessary on the behavioral surface.

Our main conclusion is that linear–cumulative thinking concerning people's future behavior is inappropriate when attempting to implement a MS model. Support and encouragement at the beginning of a series of interfaces should not automatically be extrapolated to assume such behavior throughout all of the interfaces. In addition, abrupt changes in behavior, although termed 'catastrophes', are not necessarily disasters, nor are they illogical. An abrupt change should be viewed simply as a change in perceived benefits which can be investigated and remedied as a part of the management scientist's task.

The implications of this discussion would seem to suggest that a successful management scientist is a person who:

a. views himself primarily as a change agent rather than as a technical expert;

b. takes extreme care concerning interaction with all levels of management and does not rely solely on top management support to pull him through;

c. expects abrupt changes in behavior and is ready to deal with them as logical reactions to

changes in perceptions rather than as the illogical traits of "non-scientific" managers;

d. views the task of implementation as the entire series of interactions with people rather than depending on the "logic of the model" to be the coup d'etat that suddenly inspires people to implement management science.

ACKNOWLEDGEMENT

The authors would like to acknowledge the helpful comments of Prafulla Joglekar who reviewed an earlier draft of this paper.

REFERENCES

1. EILON S (1974) Mathematical modelling for management. *Interfaces* **4**(2), 32–38.
2. GRAHAM RJ (1978) On management science process. *Interfaces* **8**(2), 80–85.
3. GRAHAM RJ (1977) The first step to successful implementation of management science. *Columbia Jl of World Bus.* **12**(3), 66–72.
4. GUPTA JND (1977) Management science implementation: experiences of a practicing OR manager. *Interfaces* **7**(3), 84–90.
5. MARKLAND RE & NEWETT RJ (1972) A subjective taxonomy for evaluating the stages of management science research. *Interfaces* **2**(2), 31–39.
6. RADNOR M & BEAN AS (1972) Top management support for management science. *Omega* **2**(1), 63–75.
7. SCHULTZ RL & SELVIN DP (Eds) (1975) *Implementing Operations Research/Management Science.* American Elsevier, New York.
8. SCHULTZ RL & SELVIN DP (1977) *An Innovation Process Perspective of Implementation.* Paper No. 601, Krannert Graduate School of Management, Purdue University, USA.
9. TURBAN E & MEREDITH J (1977) *Fundamentals of Management Science.* Business Publications, Dallas, Texas, USA.
10. URBEN GL (1974) Building model for decision makers. *Interfaces* **4**(3), 1–11.
11. ZEEMAN EC (1976) Catastrophe theory. *Sci. Am.* **234**(4), 65–83.

ADDRESS FOR CORRESPONDENCE: *RJ Graham and J Seltzer Esq, School of Business Administration, La Salle College, Philadelphia, PA* 19141, *USA.*

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 67–74
© Pergamon Press Ltd. 1979. Printed in Great Britain

0306-0483/79/0201-0067$02.00/0

Informational Impediments to the Use of Capital Budgeting Models

LAWRENCE A GORDON
The University of Kansas, USA

DAVID F LARCKER
Northwestern University, USA

FRANCIS D TUGGLE
Rice University, USA

1.7
79

(*Received August* 1978)

The primary purpose of this study was to empirically examine whether or not informational impediments impair or alter decision making, with specific focus on capital budgeting decisions. Based on detailed interviews of senior corporate executives, it appears that management information systems (MIS) are less than adequate in providing the information items required to utilize the non-naive methods of capital budgeting. It is therefore suggested that the MIS concept be vastly broadened in regard to such decisions.

INTRODUCTION

A MANAGEMENT information system (MIS) should be designed with the organization's decision making processes in mind. This point has been widely noted in the MIS oriented literature [2, 5, 9]. Yet, despite the ubiquitous backing given to the decision support MIS concept, implementation of such an idea often falls far short of the desired goal. In many such cases the result is an ineffective MIS and impaired or altered managerial decision making. There are, of course, many potential reasons behind impaired or altered decision making, of which a deficient MIS is only one. However, given the importance of MIS design to managerial decision making, MIS deficiencies must be uncovered [7, 8]. Thus, the primary objective of this paper is to empirically examine whether or not informational impediments impair or alter decision making, with specific focus on capital budgeting (CB) decisions. The issues discussed within this paper should be of key concern to those interested in MIS design as well as those concerned with the actual managerial process of CB.

There have been several empirical studies on the use of various capital budgeting methods. Recent studies on this subject have indicated that an increasingly larger percentage of firms are using discounted cash flow methods, often coupled with the naive CB methods such the payback period and the accounting rate of return [3, 10]. The more sophisticated techniques, those which go beyond the basic versions of the discounted cash flow models such as Linear Programming, PERT, and Simulations, are also gaining support among managers [6]. However, while previous studies have indicated that the percentage of firms using the non-naive methods (i.e. discounted cash flow models, as well as the more sophisticated models such as L.P.) has increased sharply, they have not examined the degree to which these techniques are being used within a particular firm. Therefore, it is not clear

whether these methods are being used on a regular or infrequent basis. As a result, these earlier studies have not been able to fully assess the impediments, if any, to using the non-naive CB methods. Our study, on the other hand, centers on examining what, if any, informational impediments exist regarding the use of non-naive CB techniques within specific firms.

In the next section we briefly discuss the main points behind the commonly discussed CB methods. That discussion is followed by an examination of the potential impediments to using the non-naive CB methods. Next we discuss the results of an empirical study which was conducted in order to assess whether or not any of the potential impediments do in fact exist, and if so, to what degree. The final section offers some concluding comments.

CAPITAL BUDGETING DECISIONS

Capital budgeting decisions may be broadly defined as those decisions which involve expenditures in projects which are expected to generate returns over several years. A firm's capital expenditures may be contrasted to its operating expenditures, where the latter are presumed to generate returns over only the current year. Thus, the time frame for capital budgeting decisions (a duration of several years is usually presumed) is fundamental in separating such decisions from the operating decisions of a firm.

Two of the simplest, sometimes referred to as naive, methods for dealing with capital budgeting decisions are the payback period and the accrual accounting rate of return. The basic version of the payback period estimates the number of years it takes to recover the cost of a project, ignoring both the time value of money and any cash flows subsequent to the payback period. The accrual accounting rate-of-return method consists of computing the ratio of the anticipated increase in average annual net income divided by either the initial increase in required investment, or half such increase, as a result of a project under consideration.

The fundamental criticism of the above two methods of capital budgeting is that they do not consider the time value of money. That is, the fact that a dollar today is worth more than a dollar one or more years from now should presumably be a factor in analyzing capital expenditures. Therefore, the present value (PV) and internal rate of return (IRR) methods of capital budgeting are often advocated in lieu of the above naive methods. Since both the PV and IRR models are highly related, we will only review the former.

The PV method of capital budgeting consists of comparing the present value of a potential project with the cost of the project, where:

$$PV = \sum_{i=1}^{n} \frac{R_i}{(1 + K)^i} \qquad (1)$$

and, $R_i = i^{\text{th}}$ period net cash flow from the project, K = the discount rate applied to the project, and n = the economic life of the project. In regards to Equation 1 above, it is often assumed that capital rationing is not an issue. Therefore, each project is considered on its own merits such that if the PV of the project is greater than its cost, the project is accepted. If, on the other hand, the PV is less than the cost, the project is rejected. Where the PV of a project and its cost are equal, the firm is presumably indifferent to the investment opportunity.

The more sophisticated capital budgeting methods all utilize some type of net cash flow discounting similar to that described above. For example, a slightly modified version of the basic discounted cash flow methods entails simulation techniques to handle the risk inherent in a project. In this regard, Hertz [4] discussed a simulation procedure whereby the individual factors affecting net cash flows are considered on an individual basis and then combined in such a fashion as to generate a probability distribution for the net cash flows. Ultimately, a portrayal of the project's investment risk was provided.

The assumption that capital rationing is not an issue is, of course, unrealistic for most firms. Instead, the usual case is for firms to have access to only a finite amount of financial resources and thus the rationing of such funds to competing projects must ensue. In this latter case, i.e. the case of capital rationing, the appropriate discount rate must be considered in the light of the alternative projects. In its basic form, such a problem has been set up as a linear programming (LP) model [11], where the objective function is to maximize the net financial assets accumulated at the horizon

Vol. 7
1979

plus the residual value of all physical assets acquired up to the horizon. Under restrictive assumptions, this objective function turns out to be the equivalent of maximizing the firm's present value. The constraints associated with such an LP model are related to the ceiling of financial resources available to the firm.

The above brief review points out that additional information is needed for the use of the non-naive capital budgeting techniques, which is not required for the naive methods. More specifically, information is required to facilitate estimating: the entire economic life of a project, the net cash flows over the entire life of the project, and the project's discount rate. Furthermore, the more sophisticated the capital budgeting technique, the greater the need for additional technical skills in order to implement the method. Accordingly, there are several *a priori* impediments to the frequent use by a firm of the non-naive CB models. These reasons are elaborated upon in the next section.

IMPEDIMENTS TO USING NON-NAIVE CAPITAL BUDGETING MODELS

As noted above, there are *a priori* impediments to the frequent use of non-naive capital budgeting models. Four such classes of reasons are (1) emotional, (2) political, (3) technical, and (4) informational. We will tersely sketch the first three, although the primary emphasis of our remarks shall focus on problems with the information set.

Where naive methods are presently in use, emotional impediments to using non-naive CB models include inertia and resistance to change, which in turn stem from the psychological comfort provided by familiar procedures and the psychological threats and dangers perceived with new, novel, and different operations. Furthermore, if the procedure currently in use is identified with its innovator, he may have additional psychological reasons for resisting the discarding of his "pet project" or "his baby".

Political impediments to using non-naive CB methods include the fear of loss of power, of the transfer of discretionary prerogatives, and of a resulting departmental or divisional status inferior to the current one. If decisions are made more scientifically, the amount of discretion may decrease, the amount of control and monitoring exerted over the operation can increase, and power naturally accrues to the ones who instituted the new decision procedures.

Technical impediments to using non-naive capital budgeting models include the fact that such models are often not known, understood, or believed in, and that the conditions of their use are not well specified. Furthermore, the models may have undesirable technical side conditions—such as, requiring special hardware (e.g. a dedicated computer), necessitating hiring special experts "to care for and feed" the models with skills not transferable elsewhere in the organization, or special algorithmic difficulties (nonconstructive solutions, slow rate of convergence, etc.).

Finally, eight different *informational impediments* to the use of non-naive capital budgeting models can readily be identified. These eight impediments are discussed below.

(1) The data for the non-naive models may not now be gathered by the organization's management information system (MIS). If a capital budgeting model requires external, and/or ex ante information (e.g. an estimation of future cash flows) and if such data are not available, the model cannot be used.

(2) The data may not be gatherable in principle. If non-naive models require data on competitors' cost functions, on future actions of governmental agencies, or otherwise requires an oracle to supply data, such models may also fail to receive us.

(3) The "wrong" data are often gathered, i.e. data irrelevant to the execution of non-naive models are often obtained. This may arise because the models are not well understood or the "right" data are not available (see points (1) and (2), above, and the next point).

(4) The data gathered may be in error. Consistency and reliability checks of the data may reveal that it has been subjected to systematic or random error processes. Such processes can be of human (perceptual biases, attitudinal biases, etc.) or mechanical (truncation, breakdowns, etc.) origin.

(5) The costs of gathering the data may exceed the benefits to be obtained from it. Often, a non-naive capital budgeting model will require data different from what is currently available. An analysis may reveal that the incremental benefits to be obtained from moving from a naive model to a non-naive one may be outweighed by the incremental costs associated with gathering the requisite data.

(6) Too much data may be required and too little time may be available to gather and process it. If non-naive CB models require data which take excessive time to gather, then the data requirements may prohibit the model from being used.

(7) The data may be unusable. In many cases, data are available only in the wrong form or format. A non-naive model may require disaggregated data, when all that is available is aggregated data. Machine readable data may be needed and only handwritten data obtainable. Quantitative data may be called for, yet only qualitative data supplied.

(8) The data may travel over unreliable information channels. The required data may be ephemeral in the sense that they manifest themselves and then depart, with no permanent record to document their fleeting existence. Examples would include reports from industrial spies, or data supplied solely by oral telephone calls with no follow-up supporting documentation.

In order to assess the existence of the above potential impediments to using non-naive CB methods, we conducted an empirical study. In the study, we primarily focus on two issues. First, how often are non-naive capital budgeting models being used within specific firms? Second, what are the informational impediments, if any, to more frequent use of such models? However, in this preliminary study, one should *not* expect to find all eight possible informational impediments discussed above corroborated. Instead, it is our intent to identify the most important and most prevalent informational impediments.

EMPIRICAL STUDY

Sample and data collection procedures

Given the difference in focus of our study from previous ones on the use of capital budgeting methods, our research methodology also differed. That is, whereas most earlier work on the use of CB techniques has been based on large scale questionnaire studies, our approach was based on in-depth interviews with groups of senior executives in charge of the capital budgeting decisions for five major U.S. corporations. This approach was chosen because it was felt that an in-depth study of a few firms would provide more meaningful data on the frequency of firm use of non-naive CB methods, and on informational impediments associated with such methods, than a large scale mail survey. Of course, what we gain from such an approach in terms of a fuller or richer understanding of a few firms we lose in terms of generalizability. On balance, however, we believe that such an approach will provide a meaningful compliment to the earlier studies on the use of various CB methods.

Prior to the actual interviews, a letter was sent to the executives involved in the study. The letter explained the *general* topic area around which the interviews would be structured. The organizations of which these executives were members were either headquarters, or a major division with general responsibility for capital expenditures at their location. The average firm's sales were in the 200 million dollar range, with most of the revenues derived from sales of machinery and other capital goods. The firms were highly capital intensive, financially stable, and publicly held.

The interviews were conducted and tape recorded at the corporate location and lasted approximately one-and-one-half hours. In all cases, the discussions were with a group of executives from each firm. Included in each group was at least one senior-level executive (vice president or president) who was involved in the capital budgeting decision process of the firm under study. The interviews had four general areas of interest: (1) the general description of the capital budgeting process and its relationship to other business activities, (2) the types of capital budgeting decisions considered by the managers, (3) the sophistication of the capital budgeting models utilized by the re-

Vol. 7
1979

spondents, and (4) determination of the different information items utilized by the executives and whether the information items were provided by the organization's existing MIS.

The manner of interviewing may be classified as a focused group interview [1]. The basic idea behind this technique is that the interviewer provides direction via questions, although an attempt is made to leave the session as unconstrained as possible. In our study, the respondents were initially asked to discuss their company's capital budgeting decision process in a fairly unstructured manner. The basic process and its relationship to the other operations of the business was detailed at this point. When the respondents began describing a capital budgeting example, the interviewer attempted to guide the interviewee's discussion to cover the type of capital budgeting analysis performed, the information items utilized, and the source of such information.

Study results

The results for the executive interviews were based on an analysis of the tape recordings and the interviewer's notes. These results are detailed below.

Non-naive capital budgeting techniques. The executives interviewed generally viewed the capital budgeting process as consisting of two broad categories: (1) non-strategic capital expenditures relating to routine types of acquisitions, such as equipment replacements, and (2) strategic capital expenditures relating to less routine types of decisions, such as building new plants or expanding into new product lines. The fundamental distinction between non-strategic capital expenditures and strategic capital expenditures initially seemed to be the difference in the dollar amounts involved. Subsequent probing and analysis of this distinction revealed that the expenditure amount was only a partial explanation. Equally important distinctions were: (1) the consequences associated with strategic expenditures were long-term in nature (i.e. a duration of five years or more), whereas non-strategic capital expenditures were usually concerned with items which needed replacement every two to five years; (2) strategic decisions involved top management in a direct manner, whereas non-strategic capital expenditures were usually executed by middle-level managers under the indirect control of senior executives; and (3) strategic capital expenditures were closely tied to the firm's long-range or strategic planning function, whereas non-strategic capital expenditures were most often classified as being related to the firm's operating plans.

The methods used by the executives to analyze the two categories of capital expenditures were related, but not identical. That is, non-strategic capital expenditures were most often analyzed in terms of a payback period, and to a lesser extent the present value or internal rate of return methods. Sometimes, both the payback period and a form of the basic discounting methods were coupled whereby the payback period operated as a constraint. In general, the greater the firm's need for the project and the lower the dollar value associated with the project the less the emphasis placed on the discounting techniques. In terms of the payback period, it was often noted that projects were expected to repay themselves within approximately four years. On the other hand, strategic capital expenditures were most often analyzed in terms of a simplified discounted cash flow procedure, sometimes coupled with a payback constraint.

The first point to note in regards to our study results is that the so-called non-naive capital budgeting models, while being used in all five firms under study, were being used in the firms on a less than regular basis. Table 1 presents a clear picture of this point. Therefore, while our findings support the notion that firms are using non-naive CB methods, they raise the question of why these methods are not used with a greater degree of regularity. This question will now be addressed, with emphasis being directed towards the information set.

Reasons for not using the non-naive models. We did not find emotional, political, or technical barriers to be significant barriers to more frequent use of non-naive CB models. On the other hand, we did discover that an important reason assigned by the executives for not using the non-naive techniques more frequently was the fact that their firm's existing MIS did not include a large part of the information needed to utilize such techniques—i.e. informational impediments represent a real barrier. More to the point, information on a project's estimated future cash flows, appropriate discount rate,

TABLE 1. FREQUENCY OF USE OF CAPITAL BUDGETING MODELS

Capital budgeting model \ Firm and type of decision	A		B		C		D		E	
	Nonstra-tegic	Stra-tegic	Nonstra-tegic	Stra-tegic	Nonstra-tegic	Stra-tegic	Nonstra-tegic	Stra-tegic	Nonstra-tegic	Stra-tegic
Payback period	O	I	I	N	A	I	O	I	I	I
Accounting rate of return	N	N	I	N	I	I	N	N	I	N
Present value or internal rate of return	I	O	O	O	I	I	I	I	O	I
Present value or internal rate of return, with a payback constraint	O	O	I	A	I	A	O	A	O	A
Simulation	I	O	N	I	N	N	N	I	I	I
Math programming	N	I	N	N	N	N	N	N	N	I
Other non-naive models (e.g., PERT and utility theory)	N	N	N	I	N	N	N	N	N	I

A = Always used; O = Often Used; I = Used on an Infrequent Basis; N = Never Used.

TABLE 2. THE IMPORTANCE ATTACHED TO THE INFORMATIONAL IMPEDIMENTS ON THE USE OF NON-NAIVE CAPITAL BUDGETING MODELS

Informational impediment \ Firm	A	B	C	D	E
(a) Data absent	L	L	L	L	L
(b) Data not gatherable	S	M	S	S	S
(c) Wrong data gathered	S	M	S	S	M
(d) Data in error	L	S	L	M	M
(e) Costs > Benefits	S	M	M	S	S
(f) Too little time	M	S	S	M	S
(g) Unusable data	M	M	M	S	S
(h) Unreliable information channels	L	M	M	L	M

Legend: S = Slight impediment—usually a minor barrier to the adoption of non-naive techniques.

M = Moderate impediment—a barrier of varying degree, depending upon the surrounding conditions, to the adoption of non-naive techniques.

L = Large impediment—usually a major barrier to the adoption of non-naive techniques.

economic life, total cost, external political factors surrounding a project's acceptance, and competitor reactions were not usually included as part of the firm's formal information system. As a result, such information had to be generated through informal, often quite subjective, inconsistent, and unreliable channels. Since the non-naive models give the appearance of exactness, it was the consensus of the executives involved in our study that the use of such models often given a false sense of 'scientific security'. Table 2 provides a detailed analysis of the informational impediments.

Other reasons often cited for not using the non-naive techniques included the non-quantifiability aspects of capital budgeting decisions, and to a much lesser extent, the limited knowledge of how to use the sophisticated techniques (see Table 2). It is worth emphasizing, however, that these two reasons were not considered significant barriers. Apparently, the wide dissemination of the non-naive techniques through universities, executive development programs, etc. have reached the intended users. Equally apparent is the fact that knowing about a particular technique is not enough to assure its applicability.

CONCLUDING COMMENTS

There are numerous articles suggesting that non-naive capital budgeting models be adopted in place of the naive analyses often employed. Recent studies suggest that such a transition has occurred. However, these earlier studies have not examined the degree to which firms are using such non-naive CB methods. We have empirically investigated two questions related to this latter point. First, to what degree are the non-naive capital budgeting techniques being used within specific firms? Second, what are the impediments, if any, to more frequent use of such models?

Based on detailed interviews of senior corporate executives, it appears that while firms claim to be using the non-naive methods of CB, these methods are used much less frequently within a particular firm than most recent studies seem to indicate. One important explanation for this fact seems to be that most formal MIS do not provide the information items required by senior executives to utilize the simplest discounted cash flow concepts, much less the more sophisticated models. Therefore, it appears that one important step required for managers to more frequently utilize the non-naive techniques for capital budgeting decisions is a broader and more comprehensively developed MIS.

REFERENCES

1. BOUCHARD TJ (1976) Field research methods: interviewing questionnaires, participant observation, systematic observation, and unobtrusive measures, in *Handbook of Industrial and Organizational Psychology* (Ed DUNNETT MD), Rand McNally, Chicago, pp 363–413.
2. DAVIS G (1974) *Management Information Systems: Conceptual Foundations, Structure and Development*, McGraw-Hill, New York.
3. FREMGREN JM (1973) Capital budgeting practices: a survey. *Mgmt Acctg* **54**(11), 19–25.

4. HERTZ DB (1964) Risk analysis in capital investment. *Harv. Bus. Rev.* **42**(1), 95–106.
5. KING W & CLELAND D (1975) The design of management information systems: an information analysis approach. *Mgmt Sci.* **22**(3), 286–297.
6. KLAMMER T (1972) Empirical evidence of the adoption of sophisticated capital budgeting techniques. *Jl Bus.* **45**(3), 387–394.
7. LUCAS H (1974) *Why Information Systems Fail.* Columbia University Press, New York.
8. MINTZBERG H (1975) *Impediments to the Use of Management Information.* Society of Management Accountants, New York.
9. MURDICK R & ROSS J (1977) *Introduction to Management Information Systems.* Prentice-Hall, Englewood Cliffs, NJ.
10. SCHALL L, SUNDEM G & GEIJSBEEK W (1978) Survey and analysis of capital budgeting methods. *Jl Fin.* **33**(1), 281–287.
11. WEINGARTNER HM (1963) *Mathematical Programming and the Analysis of Capital Budgeting Problems.* Prentice-Hall, Englewood Cliffs, NJ.

ADDRESS FOR CORRESPONDENCE: *Professor LA Gordon, The University of Kansas School of Business, Lawrence, Kansas* 66044, *USA.*

Vol. 7
1979

MEMORANDA

1.7
79

BLANK
PAGE

Vol. 7
1979

OMEGA, The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 77–79
© Pergamon Press Ltd. 1979. Printed in Great Britain
0306-0483/79/0201-0077$02.00/0

Capital Rationing and Discounting in Public Sector Investment Decisions

FOR AT least the last two decades, the discounting of future costs has been a standard analytical tool for comparing program alternatives in the public sector. Today the issue is not whether discounting is an appropriate tool for public sector investment analysis, but rather how it can be appropriately implemented. In this context, several questions arise. Is the commonly used 10 per cent discount rate correct? A Department of Defense Instruction (7041.3, October 1972) argues that it is, since a 10 per cent rate 'reflects the preference for present and future money sacrifices that the public exhibits in non-government transactions' and also since it 'represents an estimate of the average rate of return on private investment before corporate taxes and after adjusting for inflation.' Others, for example Arrow and Kurz [1] and Bradford [3], have argued for a relatively low discount rate corresponding to the 'marginal rate of social time preference.' Further, if one is discounting only streams of costs, how can alternatives which make differing contributions to overall system effectiveness be compared? Does this mean that discounting should be limited to the trade-off of exactly substitutable alternatives?

In this memorandum we suggest answers to these questions on a limited scale. We do not take the perspective of decision-makers such as the Office of Management and Budget (OMB), who must make trade-offs among programs which are generically quite different. Such program decisions generally have strategic implications for the agencies they impact, frequently altering significantly the agency's budget levels and in some instances even establishing new agencies. We address ourselves instead to those who must make tactical decisions for established programs, working within essentially fixed budgets to achieve specified minimum levels of overall effectiveness from now to some horizon. These less strategic decision makers are faced with a variety of alternatives, project, or systems which are complementary and partially substitutable in meeting these minimum requirements.

The objective of this note is to develop a set of discount rates that depend upon the optimal configuration of projects which meets the requirements. These discount rates then become part of a 'modified' present value calculation that includes both project costs and 'imputed' benefits for contributing to system effectiveness during constraining periods. The modified present value calculation can be used by the decision maker to (i) communicate those projects which are part of the optimal plan; (ii) evaluate a new project in relation to the current optimal plan with which it must compete for funds; and (iii) identify the need for new projects which offer increased effectiveness in constraining periods and avoid using funds during periods when the budgets are being used to great advantage. The proposed model and the resulting implications represent an extension to the public sector of the private sector analysis in Bradley and Frey [4].

A PUBLIC SECTOR MODEL

An investment decision model for the public sector manager is not a straightforward application of the private sector models of [2], [4], and [5]. Certainly the public sector investments, programs, or systems involve streams of cash flows, and the discounting off these cash flows is a common form of analysis. However, what appears to make the analysis in the private sector 'go through' is that the returns from an investment are measured in the same units as the investment itself, i.e. dollars. This is clearly not the case in the setting in which public sector investment problems naturally arise. The usual condition is that the investments are streams of dollar costs; however, the returns are measured not in dollars, but in some more relevant measure of effectiveness. In many agencies, the criteria for measuring the effectiveness of a program are quantifiable and well established.

In those situations where effectiveness can be quantified and measured and a single criterion is of primary importance, a public sector analogue of the private sector model in [4] can be formulated. Its interpretation has several interesting aspects. If we assume for the moment that the planning horizon for the particular agency or department under consideration is continuously being revised so that it remains a constant number of years into the future, it might be reasonable to assume that an appropraite objective function is to maximize the horizon effectiveness subject to meeting minimum performance requirements in the intervening years[1] and satisfying budgetary limitations. The *horizon effectiveness* model can then be formally stated as:

$$\begin{array}{lll} \text{Maximize } r & \text{Shadow} & \\ \text{subject to:} & \text{prices} & \\ \mathbf{Cx} \leqq \mathbf{b} & :\phi & \\ \mathbf{Ex} - u_n r \geqq \mathbf{r} & :-\lambda & (1) \\ 0 \leqq \mathbf{x} \leqq \mathbf{m} & & \end{array}$$

where r is the horizon value of total effectiveness, c_{ij} is the cost of project j in period i (note that $c_{ij} > 0$ is assumed to be a cash outflow), x_j is the level of investment in project j, b_i is the total budget available for all projects in period i, e_{ij} is the contribution to total effectiveness by project j in period Hi, r_i is the required level of effectiveness for the department or agency in period i, and m_j is the maximum limit for investment in project j. It has been assumed that budget surpluses are not carried over from one period to the next; however, if in a particular application surpluses can be carried over, 'dummy' projects can be included in the model which contribute nothing to effectiveness but have a unit cash outflow at time i and a unit cash inflow at time $i + 1$.

[1] Even though a single measurement of effectiveness may be used in the objective function, several sets of effectiveness constraints may be imposed. In this way, the multiple criteria aspect of many situations can be partially modeled.

The important point to recognize in relating the shadow prices of the horizon effectiveness model to discount rates is that discount rates do *not* themselves say anything about the *units* in which the criterion is measured. The one-period discount factor β_i merely states that one dollar received at the end of period i is equivalent to β_i dollars at the beginning of period i. As such, the one-period discount factor β_i is an indifference relationship between receiving budgets at two distinct periods in time. The indifference is with regard to the maximization of the criterion but is independent of the units in which the criterion is measured. Hence we can argue that discount rates naturally arise in the public sector model (1) just as they do in the private sector model of [4]. In addition, the same relationship between the one-period discount factors and the optimal shadow prices hold for both the private and the public sector models.

$$\beta_i = \frac{\phi_i}{\phi_i - 1}.$$

The many-period discount factor is then

$$\rho_i = \prod_{k=1}^{i} \beta_k = \frac{\phi_i}{\phi_0} \qquad i = 0, 1, \ldots, n.$$

IMPLICATIONS FOR DECISION AND CONTROL

In the private sector model we were able to establish an equivalence between the linear programming reduced costs for a project and the net present value of that project using the imputed discount factors. These discount factors could then be used in the decision and control processes of the division. In order to be able to make a similar interpretation of the reduced costs of the public model (1), we must first interpret the shadow price $-\lambda_i$ associated with the effectiveness in period i constraints. The sign was chosen so that the optimal shadow prices satisfy $\boldsymbol{\lambda} \geqq \mathbf{0}$. λ_i is simply the increase (decrease) in horizon effectiveness resulting from a unit *decrease* (increase) in required effectiveness in interim year i. Hence λ_i is the exchange rate between horizon effectiveness and interim effectiveness. It states the worth, or 'price', of a unit of interim effectiveness in year i in terms of a unit of horizon effectiveness.

The reduced costs of the horizon effectiveness model have the following form and implications:

$$\boldsymbol{\lambda}\mathbf{E}_j - \boldsymbol{\phi}\mathbf{C}_j \begin{cases} > 0 \rightarrow x_j = m_j \text{ (project accepted);} \\ < 0 \rightarrow x_j = 0 \text{ (project rejected);} \\ = 0 \rightarrow 0 \leqq x_j \leqq m_j \text{ (project marginal).} \end{cases} \tag{2}$$

Since v must be a basic variable in (1), its reduced cost is $1 - \lambda_n = 0$; hence $\lambda_n = 1$. The interpretation of the reduced cost is made easier if we rewrite (2) specifically breaking out the criterion (horizon effectiveness) as

$$e_{Nj} + \boldsymbol{\lambda}'\mathbf{E}'_j - \boldsymbol{\phi}\mathbf{C}_j \tag{3}$$

In (3), $\boldsymbol{\lambda}'$ and $\mathbf{E}'_j$ are identical to and include all but the last terms of $\boldsymbol{\lambda}$ and $\mathbf{E}_j$, respectively. We can now interpret the terms of (3) as follows: For project j, e_{Nj} is the level of effectiveness at the horizon, $\boldsymbol{\lambda}'\mathbf{E}'_j$ is the value of interim effectiveness, and $\mathbf{C}_j$ is the opportunity cost of investing in the program or system.

To interpret the reduced costs of equation (3) in terms of the discounted present value of the cost stream, we divide (3) by ϕ_0, which gives

$$\frac{1}{\phi_0}(e_{Nj} + \boldsymbol{\lambda}'\mathbf{E}'_j) - \frac{\phi}{\phi_0}\mathbf{C}_j. \tag{4}$$

Expression (4) has the same sign as (3) since the initial budgetary constraint is assumed binding, implying that $\phi_0 > 0$. Recognizing that the derived discount factors are just $\boldsymbol{\rho} = \pi/\phi_0$ and defining $\rho_0 \equiv 1/\phi_0$, we have

$$\rho_0(e_{Nj} + \boldsymbol{\lambda}'\mathbf{E}'_j) - \boldsymbol{\rho}\mathbf{C}_j \tag{5}$$

We can now interpret (5) in an interesting way. The term $\boldsymbol{\rho}\mathbf{C}_j$ is simply the discounted present value of the cost stream $\mathbf{C}_j$ associated with a unit level of investment in project j. The discount factors are those imputed from the public sector model (1). The term

$$(e_{Nj} + \boldsymbol{\lambda}'\mathbf{E}'_j)$$

is the overall effectiveness of project j, consisting of the value of its horizon effectiveness e_{Nj} and the value of its interim effectiveness $\boldsymbol{\lambda}'\mathbf{E}'_j$. The shadow prices $\boldsymbol{\lambda}'$ describe the exchange rates between the two types of effectiveness, interim and horizon. The factor $\rho_0 \equiv 1/\phi_0$ is measured in units of dollars of initial budget per unit of horizon effectiveness; hence ρ_0 has the interpretation of present worth, or current 'price' of a unit of horizon effectiveness. Expression (5) then merely compares the dollar worth of horizon effectiveness (current price times the amount of horizon effectiveness) with the discounted cost of attaining that worth by investing in project j.

We can now establish a decision and control system based on the mechanism given by (5). During the formal budgeting process, the public sector linear programming model given in (1) is solved, and a collection of projects turns out to be optimal. Along with the optimal solution, the parameters of a control system are also determined; these parameters are $\boldsymbol{\rho}$, $\boldsymbol{\lambda}$, and ρ_0. The vector $\boldsymbol{\rho}$ is the vector of discount factors for determining the present value of the cost stream associated with any project. The vector $\boldsymbol{\rho}$ is the vector of exchange rates that fully describe the trade-offs between various levels of interim effectiveness and horizon effectiveness. Note that $\boldsymbol{\lambda}$ gives measures of the relative strategic importance of investing in programs or systems that contribute to required effectiveness in years before the horizon. Finally, ρ_0 is the imputed current dollar worth of a unit of horizon effectiveness. Hence, the effectiveness of a program or system, both horizon and interim, is valued at ρ_0 and compared to the discounted present value of the cost stream for that project. Programs are then accepted or rejected according to:

$$\rho_0(e_{Nj} + \boldsymbol{\lambda}'E'_j) - \rho C_j \begin{cases} > 0 \rightarrow x_j = m_j \\ \quad \text{(project accepted)} \\ < 0 \rightarrow x_j = 0 \\ \quad \text{(project rejected);} \\ = 0 \rightarrow 0 \leqq x_j \leqq m_j \\ \quad \text{(project marginal).} \end{cases} \tag{6}$$

The marginal projects must be selected so as to satisfy the budget and effectiveness constraints, taking the accepted projects into account. The control system then can communicate those programs or systems accepted and rejected without appealing to the linear programming model directly. Between budgeting cycles, new programs or systems will arise that were not originally analyzed. The programs should be accepted or rejected also according to (6), with revisions to the parameters of the control system as suggested in [4].

We do not claim to have solved the 'discounting problem' for the public sector. However, we hope that,

Vol. 7
1979

at least for those situations where the investments can be thought of as being continuous and effectiveness can be quantified and measured, we have successfully argued that discount rates for tactical decisions arise in a natural way without appealing to external and often irrelevant forces. Further, a decision and control system using discounting concepts results from considering an undiscounted linear programming formulation that maximizes horizon effectiveness, or some other appropriate measure of performance, subject to budgetary and interim effectiveness constraints.

REFERENCES

1. ARROW KJ & KURZ M (1970) *Public Investment, the Rate of Return and Optimal Fiscal Policy*. Johns Hopkins Press, Baltimore, MD.
2. BAUMOL WJ & QUANDT RE (1965) Investment and discount rate under capital rationing—a programming approach. *Econ. Jl* **75**(298), 317–329.
3. BRADFORD DF (1975) Constraints on government investment opportunities and the choice of discount rate. *Am. Econ. Rev.* **65**(5), 887–899.
4. BRADLEY SP & FREY SC (1978 Equivalent mathematical programming models of capital rationing. *Jl Fin. Quant. Anal.* **13**(2), 345–361.
5. WEINGARTNER HM (1963) *Mathematical Programming and the Analysis of Capital Budgeting Problems*. Prentice–Hall, Englewood Cliffs, NJ.

Stephen P Bradley
Sherwood C Frey Jr

Harvard University
Graduate School of Business Administration
Soldiers Field
Boston
MA 02163
USA

(*June* 1978)

OMEGA, The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 80–81
© Pergamon Press Ltd. 1979. Printed in Great Britain
0306-0483/79/0201-0080$02.00/0

A Decision Model with Stochastic Cost and Demand Curves

THE STANDARD microeconomic textbooks consider pricing decisions for profit maximization when the selling price-market demand relationship and the production volume–production cost relationship are given deterministically; and in recent research efforts (e.g. [2] and [3]), the probability distribution of a product's profit is derived from stochastic estimates of market demand, market price and production cost. In the first model 'price' is the decision variable and the 'world' is completely deterministic; in the second model the stochastic nature of the 'world' is explicitly recognized but the effects of managerial decisions on cost and sales figures are completely ignored. This paper attempts to combine the favorable characteristics of both models and considers the realistic situation in which the firm has to set a selling price for the product; therefore, 'price' is not a stochastic variable but a decision variable. At the same time, the price–demand and volume–cost relationships are known only in stochastic terms. The problem is then to determine the selling price to optimize the firm's objective.

LINEAR DEMAND AND COST CURVES

Assume that, through a combination of statistical analysis of historical data and experts' forecasts, the stochastic relationship between unit selling price P and the planning period's market demand D for a given product is given by the regressive linear function

$$D = a_0 + a_1 P + \alpha \tag{1}$$

and that the total cost T of manufacturing D units during the planning period is estimated to be

$$T = b_0 + b_1 D + \beta, \tag{2}$$

where α and β are normally distributed error terms with mean 0 and standard deviation σ_α and σ_β respectively. Assuming that the plant can always manufacture sufficient units to meet the actual demand in the current period, the period's profit Z from the product is then

$$\begin{aligned} Z &= DP - T \\ &= (a_0 + a_1 P + \alpha)P - b_0 \\ &\quad - b_1 (a_0 + a_1 P + \alpha) - \beta \end{aligned} \tag{3}$$

Z is clearly stochastic, with mean and standard deviation as

$$\begin{aligned} E(Z) = &-a_1 P^2 + (a_0 + a_1 b_1)P \\ &- (b_0 + a_0 b_1) \end{aligned} \tag{4}$$

and

$$\sigma(Z) = (P^2 \sigma_\alpha^2 + b_1^2 \sigma_\alpha^2 + \sigma_\beta^2)^{1/2} \tag{5}$$

If the objective is to maximize expected profit, then by solving $\mathrm{d}[E(Z)]/\mathrm{d}P = 0$, it can be easily verified that to set the selling price at

$$P_E^* = (a_0 - a_1 b_1)/2a_1 \tag{6}$$

will maximize $E(Z)$. However, it is now well recognized that, under condition of uncertainty, a decision maker should consider the trade-off between the expected return and the risk (or the return's standard deviation) resulting from an action. One common way to effect this trade-off explicitly is to maximize the utility function

$$U(Z) = E(Z) - k\sigma(Z), \tag{7}$$

where k is the mean-standard deviation trade-off factor. The value of P_U^* that maximizes $U(Z)$ cannot be expressed in a convenient algebraic form, but lower and upper bounds of P_U^* can be easily derived. Figure 1 sketches the curves $E(Z)$, $\sigma(Z)$ and a straight line $kP\sigma_\alpha$. From (7), $[\mathrm{d}E(Z)/\mathrm{d}P]_{P_U^*} = [\mathrm{d}\sigma(Z)/\mathrm{d}P]_{P_U^*}$; i.e. P_U^* is the P-value where $E(Z)$'s gradient equates $\sigma(Z)$'s gradient, and Fig. 1 illustrates clearly that $P_U^* < P_E^*$ always holds. The point

$$P_1 = (a_0 + a_1 b_1 - k\sigma)/2a_1 \tag{8}$$

is the point where $E(Z)$'s gradient equates the gradient of the line $kP\sigma_\alpha$, and Fig. 1 also illustrates that $P_U^* > P_1$ always holds. Hence P_U^* are bounded as

$$P_E^* > P_U^* > P_1 \tag{9}$$

and an accurate value of P_U^* can be found numerically between the bounds by using standard search techniques (see, e.g. [4]).

As a numerical example, let $k = 1$, and

$$\left.\begin{aligned} D &= 1000 - P + \alpha, & \sigma_\alpha^2 &= 5 \times 10^3 \\ T &= 5000 + 100D + \beta, & \sigma_\beta^2 &= 16 \times 10^6 \end{aligned}\right\} \tag{10}$$

From the foregoing, it follows that $P_E^* = 550$, $P_1 = 514.6$, and P_U^* that maximizes $U(Z)$ is somewhere in between. Using the Fibonacci Search subroutine in the IMSL library [1], P_U^* is found to be \$515.5, with corresponding $E(Z) = 196{,}310$ and $\sigma(Z) = 37{,}346$. Given the simplicity of (7), the same answers can also be obtained in about ten minutes using a calculator.

NON-LINEAR DEMAND AND COST CURVES

Although it is well recognized that the price–demand and volume–cost relationships are often non-linear, the linear relationships considered above are really quite adequate for most cases. This is because the accuracy of the relevant results depends only on the accuracy of the functional relationships in the neighborhood of P_U^*. Thus, if necessary, one can obtain a crude initial estimate of P_U^* using linear functions that approximate the non-linear relationships over a wide range of P values; after that, refined linear functions closely fitted to the non-linear relationships in the neighborhood of P_U^* can be used to obtain more accurate estimates of P_U^* and other relevant values.

Nevertheless, if one insists on using non-linear demand and cost functions, the approach of the preceding section can be easily extended to handle the higher degree of polynomials of the demand-cost functions, even though the algebraic manipulations will become more tedious. For example, for the case of quadratic demand and cost curves

$$\left.\begin{aligned} &D = a_0 + a_1 P + a_2 P^2 + \alpha = D_0 + \alpha \\ &\text{and} \\ &T = b_0 + b_1 D + b_2 D^2 + \beta \end{aligned}\right\} \tag{11}$$

Vol. 7
1979

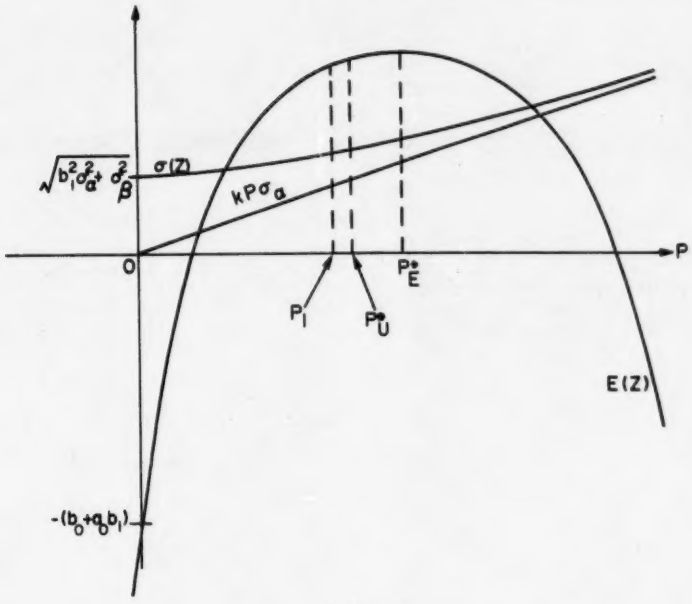


FIG. 1.

where α and β are normally distributed error terms, one can derive the mean and variance of the profit $Z = DP - T$ as

$$\left.\begin{aligned} &E(Z) = (P - b_1)\, D_0 - b_0 - b_2 D_0^2 - b_2\sigma_z^2 \\ &\text{and} \\ &\mathrm{Var}(Z) = 2b_2^2\sigma_z^4 + \sigma_\beta^2 + [2a_2b_2P^2 \\ &\qquad - (1-2a_1b_2)P + [b_1 + 2a_0b_2)]^2\sigma_z^2 \end{aligned}\right\} \quad (12)$$

Given (12), the value of P_U^* that maximizes $E(Z) - k\sigma(Z)$ can be easily determined by standard search techniques.

CONCLUSION

In a typical pricing, capacity planning or product selection problem, the management has to make a number of decisions. The effects of these decisions are only known in a stochastic sense, and the decisions should be made such that they are consistent with the firm's objective on making profit and avoiding risk. The model presented in this paper attempts to incorporate realistically all the above three components, whereas earlier models (see e.g. [2] and [3] have often overlooked (i) the stochasticity of the environment, (ii) the difference between the decision variables and the stochastic environmental factors, or (iii) the risk-aversion component in the firm's objective. Although the presented model considers only one decision variable, we should point out that the addition of other relevant variables (e.g. advertising expenses in the demand function and subcontracting in the cost function causes no conceptual difficulty. Practical solution of these models are straightforward with standard search procedures.

REFERENCES

1. International Mathematical and Statistical Libraries (1975) IMSL Library 1, Reference Manual, Edition 4 (Fortran IV) S/370/360.
2. KOTTAS J, LAU AHL & LAU HS (1978) A general approach to stochastic management planning models: an overview. *Acctg Rev.* **52**, 389–401.
3. LIAO M (1975) Model sampling: a stochastic CVP analysis. *Acctg Rev.* **49**, 780–790.

John F Kottas
Hon-Shiang Lau

(*September* 1978)

School of Business Administration
8001 *Natural Bridge Road*
University of Missouri
St Louis
Missouri 63121
USA

OMEGA, The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 82–84
© Pergamon Press Ltd, 1979. Printed in Great Britain
0306-0483/79/0201-0082$02.00/0

The Status of Variable Costs in Profitability Planning

INTRODUCTION

IN DESIGNING their profitability model for tactical planning [3], Eilon and Cosmetatos made a series of assumptions on the structure of the functional relationships between the major variables in the model. The sensitivity of the rate of return to the relationship between working capital and the level of activity in the firm was shown to be relatively small, implying that the practical problems associated with data collection and model evaluation can be simplified considerably.

When examining the case of a change in the volume triggered by external forces and the case of a change in the selling price, it was assumed that the unit variable cost remains unchanged and hence that the relative change in the unit cost simply reflects the resulting change in the unit fixed cost. This assumption is supported by many empirical studies such as the ones by Dean [1] in a hosiery mill, a leather belt shop, a hosiery department and a shoe department, but not by others. It is the purpose of this note to modify the analysis to take account of a linear relationship between the unit variable cost and output and test the results against empirical evidence.

THE EFFECTS OF A CHANGE IN THE VOLUME

Assume that the volume V changes to $V + \delta V = V(1 + V^*)$ where δV is the incremental change and V^* the relative change (i.e. $V^* = \delta V/V$). If δV is relatively small, it may be reasonable to assume that over the range $V \ldots V + \delta V$ the unit variable cost s can be expressed as a linear function of V, say, $s = k_1 V + k_2$ (this is a sufficiently broad assumption to allow the effect of increasing V to be either beneficial or detrimental to the unit cost). Following the method of incremental analysis [2] it can be shown that the change V^* will change s by

$$s^* = s_1 V^* \tag{1}$$

where $s_1 = k_1 V/s$. If we ignore the secondary effects on the unit cost c of changes in working capital requirements, then we have, following the methodology outlined in [3],

$$c^* = f_1 s_1 V^* - (1 - f_1)\frac{V^*}{1 + V^*} \tag{2}$$

where f_1 is the ratio of the unit variable cost s to the total cost c. Similarly, the effect of V^* on the rate of return r is to change it by

$$r^* \simeq -\frac{1-\alpha}{\alpha} f_1 s_1 V^*(1 + V^*) + (V^*/\alpha)[1 - f_1(1 - \alpha)] \tag{3}$$

where $\alpha = (p - c)/p$ denotes the profit margin. The evaluation of formulae (2) and (3) is presented in Table 1 for $0 \le f_1 \le 1$, $-1 \le s_1 \le 1$, $V^* = 0.01$ and $\alpha = 0.10$. The results in the Table clearly show that the values of c^* and r^* are quite sensitive to the assumption $s_1 = 0$ when the proportion of variable to total costs is large.

EMPIRICAL EVIDENCE

In order to gain an insight into the magnitude of s_1 and f_1 values which are likely to be encountered in practice, we resorted to empirical studies of short run cost–output relationships reported in the literature. When analysing the data, we were faced with three kinds of difficulties:

(1) In some of the studies, the relationship between the unit cost and the level of output is confined to a constituent element of the total cost only. Nordin [7], for example, studies the short run fuel cost in steam electricity plants and fits the following curvilinear equation to the data:

$$Y(x) = 16.68 + 0.125x + 0.00439x^2 \tag{4}$$

where

$Y(x)$ = total fuel cost

TABLE 1. VALUES OF c^* AND r^* (%) FOR $V^* = 0.01$ AND $\alpha = 0.10$

f_1	c^*, r^* \ s_1	−1.0	−0.8	−0.6	−0.4	−0.2	0	0.2	0.4	0.6	0.8	1.0
0.0	c^*	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0
	r^*	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
0.2	c^*	−1.0	−1.0	−0.9	−0.9	−0.8	−0.8	−0.8	−0.7	−0.7	−0.6	−0.6
	r^*	6.4	6.8	7.1	7.5	7.8	8.2	8.6	8.9	9.3	9.6	10.0
0.4	c^*	−1.0	−0.9	−0.8	−0.8	−0.7	−0.6	−0.5	−0.4	−0.4	−0.3	−0.2
	r^*	2.8	3.5	4.2	5.0	5.7	6.4	7.1	7.8	8.6	9.3	10.0
0.6	c^*	−1.0	−0.9	−0.8	−0.6	−0.5	−0.4	−0.3	−0.2	−0.0	0.1	0.2
	r^*	−0.8	0.3	1.4	2.4	3.5	4.6	5.7	6.8	7.8	8.9	10.0
0.8	c^*	−1.0	−0.8	−0.7	−0.5	−0.4	−0.2	−0.0	0.1	0.3	0.4	0.6
	r^*	−4.4	−3.0	−1.5	−0.0	1.4	2.8	4.2	5.7	7.1	8.6	10.0
1.0	c^*	−1.0	−0.8	−0.6	−0.4	−0.2	0.0	0.2	0.4	0.6	0.8	1.0
	r^*	−8.0	−6.2	−4.4	−2.6	−0.8	1.0	2.8	4.6	6.4	8.2	10.0

TABLE 2. EVALUATION OF FORMULA (2) FOR DATA ON A TEXTILE FIRM [4]

		$F = 50{,}000$			c^* (%)		$F = 90{,}000$			c^* (%)	
V (yards)	Total Cost ($)	f_1	s	s_1	s_1 in col. 5	$s_1 = 0$	f_1	s	s_1	s_1 in col. 10	$s_1 = 0$
1	2	3	4	5	6	7	8	9	10	11	12
75,000	148,500	0.663	1.313	−0.059	−0.37	−0.33	0.394	0.780	0.412	−0.44	−0.60
100,000	178,700	0.720	1.287	−0.099	−0.35	−0.28	0.496	0.887	0.216	−0.39	−0.50
125,000	206,875	0.758	1.255	−0.080	−0.30	−0.24	0.565	0.935	0.176	−0.33	−0.43
150,000	235,200	0.787	1.235	0.170	−0.08	−0.21	0.617	0.968	0.459	−0.10	−0.38
175,000	272,300	0.816	1.270	0.171	−0.04	−0.18	0.669	1.042	0.396	−0.06	−0.33
200,000	310,200	0.839	1.301	—	—	—	0.710	1.101	—	—	—

and

x = output as a percentage of capacity.

On the basis of (4) it is impossible to ascertain whether $s(x) = 0.125 + 0.00439x$ is the total unit variable cost of the plant or not; and it is also impossible to evaluate f_1, for the added reason that the total fixed cost is not known. However, if f_1 is assumed to represent the proportion of the unit variable fuel cost to the total unit fuel cost and s_1 is defined accordingly, then (2) can be used to assess the effect on the total unit fuel cost of changes in capacity utilisation. Thus, if say $x = 70\%$, we have $f_1 = s(70)/Y(70) \simeq 0.645$ and $s_1 = 0.00439X70/s(70) \simeq 0.711$ and the effect of a one per cent increase in the level of output ($V^* = 0.01$) would be to increase the total unit fuel cost by 0.11%. (Had the assumption $s_1 = 0$ been made instead, it would have been found that a 1% increase in the plant's utilisation reduces the total unit fuel cost by 0.35%).

(2) In many of the case studies (for example, the studies by Dean [1] in a furniture factory and a coat department, the one by Markham [6] in the rayon industry, etc.), some fixed cost items (such as depreciation and overheads) and in one case (the study by Johnston [5] in a food processing firm) all fixed costs are regarded by the authors as irrelevant to the objectives of their analysis, and their magnitude is not reported. As a result, the f_1 values calculated from the data do not correspond to the actual proportion of variable to total costs, and the implementation of formulae (2) and (3) tends to overestimate the effects on c^* and r^* of the assumption $s_1 = 0$ by an arbitrarily high or low margin of error.

(3) In most of the case studies, the cost data cover a range of output levels from, say, V_1 to V_2, where

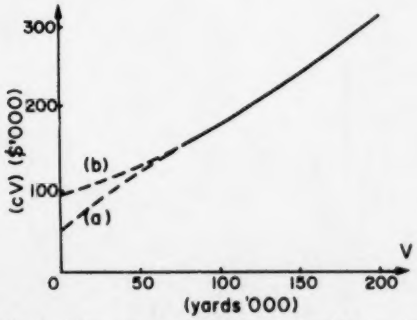


FIG. 1. *Short run cost–output relationship for data in Table 2.*

$V_1 > 0$. Thus, the cost at $V = 0$ (the fixed cost) can only be inferred by extrapolation which may lead to large errors if $V_1 \gg 0$. Consider, for example, the data on the short run cost relationship of one of the products (item 345, of Atherton Company, a large textile firm [4]) reproduced in Table 2, columns 1 and 2. If represented graphically (see Fig. 1), they indicate that the total cost is not a linear function of V and hence that s_1 is not zero over the whole range of V values (between 75 and 200 thousand yards). The solid line in Fig. 1 can be projected until $V = 0$ in many different, all arbitrary, ways. Two of them will now be considered. If curve (a) is adopted, then the total fixed cost is $50,000, and columns 3 and 4 in Table 2 give the resulting values of f_1 and s for each of the six output levels. If it is now assumed that the unit variable cost s is a linear function of V within each class of V values, then estimates for s_1 are not difficult to obtain using (1) (see Table 2, column 5). Columns 6 and 7 give respectively the effects on c of an increase in V by 1 per cent as calculated using formula (2) for the two cases where s_1 takes the values in column 5 or is assumed to be zero. For a total fixed cost of $90,000 (curve (b) in Fig. 1) the f_1 and s values are, as expected, smaller than before but, quite surprisingly, the s_1 values are now markedly different. As a result, any conclusions regarding the magnitude of s_1 and f_1 values seem to be heavily dependent on the assumptions that the analyst is prepared to make.

CONCLUSIONS

In practice, firms adopt accounting conventions for classifying costs into variable and fixed. Subject to these conventions, an estimate for f_1, the proportion of variable to total costs, is not difficult to obtain. If f_1 is large, then the effect on the unit cost and the return on capital of a change in the level of output is sensitive to the assumption that the unit variable cost remains unchanged (see Table 1); in such situations, therefore, an attempt to estimate the resultant expected relative change in the unit variable cost and the subsequent use of (2) and (3) instead of the formulae derived in reference [3] might be justified.

REFERENCES

1. DEAN J (1976) *Statistical Cost Estimation.* Indiana University Press, Bloomington.
2. EILON S (1976) Sensitivity of the rate of return. *Omega* 4(3), 350–352.

3. EILON S & COSMETATOS GP (1977) A profitability model for tractical planning. Omega **5**(6), 673–688.
4. HAYNES WH (1969) *Managerial Economics*. Business Publications, Austin, Texas.
5. JOHNSTON J (1960) *Statistical Cost Analysis*. McGraw–Hill, New York.
6. MARKHAM JW (1952) *Competition in the Rayon Industry*. Harvard University Press, Cambridge, MA.
7. NORDIN JA (1947) Note on a light plant's cost curves. *Econometrica* **15**, 231–235.

George Cosmetatos
Nicholas Barbaroussis
Samuel Eilon

(*October* 1978)

Department of Management Science
Imperial College
Exhibition Road
London SW7 2BX
UK

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 1, pp. 85–87
© Pergamon Press Ltd. 1979. Printed in Great Britain
0306-0483/79/0201-0085$02.00/0

International Register of Theses and Dissertations

The following are obtainable at the MIT Microreproduction Laboratory, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02139, USA:

ADAMS DF (with HOLTCAMP MS & MARTIN F) (1974) Engineers and Physicians: What Turns Them On? SM Thesis, MIT, Cambridge, MA.

ANDERSON LR (1974) A Performance Review of the US Atomic Energy Commission's Benefit–Cost Analysis Program for Civilian Nuclear Power Reactors. SM Thesis, MIT, Cambridge, MA.

ARMSTRONG RJ (1974) The Use of Numerically Controlled Machine Tools Aboard a Naval Tender: A Job Shop Design. SM Thesis, MIT, Cambridge, MA.

BARNES AH (1974) A Case Study of the Nationalization of the International Petroleum Company and the Cerro de Pasco Corporation in Peru. SM Thesis, MIT, Cambridge, MA.

BECK GE (1974) Office Space Design and the Designer. SM Thesis, MIT, Cambridge, MA.

BECKER RE (1974) Success and Failure of Efforts to Enter New Markets by Aerospace Organizations. SM Thesis, MIT, Cambridge, MA.

BERT RL (1974) The 1965 Reform of Taxation of Dividends in France: The Reform and Its Consequences. SM Thesis, MIT, Cambridge, MA.

BITRAN GR (1974) Fractional Programming Duality, Algorithms, Sensitivity Analysis and Applications. SM Thesis, MIT, Cambridge, MA.

BODHISOMPSOM S (1974) The Effect of Domestic and Foreign Environments on a Banking Information System. SM Thesis, MIT, Cambridge, MA.

BROGAN SJ (1974) A Look at Worker Alienation in State and Local Government. SM Thesis, MIT, Cambridge, MA.

CAMPBELL WL (1974) Control and Status Reporting in Defense Weapon System Acquisition Projects in the US Air Force. SM Thesis, MIT, Cambridge, MA.

CATANZANO LA (1974) The Dynamics of an Organization: A Case Study of the Institutions of Student Life. SM Thesis, MIT, Cambridge, MA.

CHANG DH (1974) Shift Work: Management Techniques. SM Thesis, MIT, Cambridge, MA.

CHANG LC (1974) Required Rate of Return of Public Utility Equities, 1963 to 1972. SM Thesis, MIT, Cambridge, MA.

CHARTOUNI NN (1974) A Model for the Development of the Financial Sector in a Small Economy: The Case of Lebanon. SM Thesis, MIT, Cambridge, MA.

CODY AM (1974) The Use of Citizen Surveys in Municipal Administration. SM Thesis, MIT, Cambridge, MA.

COHEN BD (1974) Use and Potential Applications of Models for Funancial Planning. SM Thesis, MIT, Cambridge, MA.

CULP WS (1974) School Health Planning for the City of Boston. SM Thesis, MIT, Cambridge, MA.

DE CHATILLON RA (1974) Energy Use in the Steel Industry. SM Thesis, MIT, Cambridge, MA.

DE LA CHAPELLE R (1974) Improving ROI Measurement of Investment Center Performance. SM Thesis, MIT, Cambridge, MA.

EINSIDLER JE (1974) The Labor Productivity Index in a High Technology Industry. SM Thesis, MIT, Cambridge, MA.

FIELDING EV (1974) Some Strategic Ownership Considerations for Foreign Investors in the Andean Pact Region. SM Thesis, MIT, Cambridge, MA.

FOSSEEN JB (1974) Linearization of the Urban Dynamics Model at Equilibrium. SM Thesis, MIT, Cambridge, MA.

FOUCART BR (1974) Aircraft Leasing: The Case of Europe. SM Thesis, MIT, Cambridge, MA.

GOLDEN BL (1974) A Minimum-Cost Multicommodity Network Flow Problem Concerning Imports and Exports. SM Thesis, MIT, Cambridge, MA.

GRADER CR (1974) The Least Developed Countries: A Useful Categorization for Aid Strategy? SM Thesis, MIT, Cambridge, MA.

GRONNING TG (1974) Corporate Policy and Social Responsiveness. SM Thesis, MIT, Cambridge, MA.

HALPHEN J-P (1974) Risk in the Oil Industry, Market Determination and Accounting Explanation. SM Thesis, MIT, Cambridge, MA.

HAWLEY HB (1974) Effective Advertising Themes Through Applied Multidimensional Scaling: A Study of the Market for Public Transportation. SM Thesis, MIT, Cambridge, MA.

HORI S (1974) Ownership Strategy in International Business Management—Japan, A Case Study. SM Thesis, MIT, Cambridge, MA.

ISSACS DL (1974) Centralization or Decentralization? A Study of the Data Processing Requirements of the Research Department of the Federal Resarve Bank of Boston. SM Thesis, MIT, Cambridge, MA.

IZUTSU S (1974) Diversification of Drug Manufacturers in the United States. SM Thesis, MIT, Cambridge, MA.

JU C (1974) A Market Study of a Solid State Power Conversion System. SM Thesis, MIT, Cambridge, MA.

JUNG WEO (1974) Foreign Labor in Germany and Switzerland. SM Thesis, MIT, Cambridge, MA.

KAPLAN PD (1974) An Investigation of the Decision Process Among Faculty and Student Readers of Applications to the Masters Program in Management. SM Thesis, MIT, Cambridge, MA.

KONIDARIS N (1974) Planning and Designing Improvements in a Management Information System: A Case Study. SM Thesis, MIT, Cambridge, MA.

LIM HS (1974) Planning in Port Authorities. SM Thesis, MIT, Cambridge, MA.

LIPSEY S (1974) A Minicomputer-Based Factory Data Collection System. SM Thesis, MIT, Cambridge, MA.

LUTES GS (1974) Project Selection and Scheduling in the Massachusetts Department of Public Works. SM Thesis, MIT, Cambridge, MA.

MAIETTA CA (1974) Stereotypes: The Interpersonal Perceptions of Undergraduate Students as Related to the Interpersonal Perceptions of R&D Personnel. SM Thesis, MIT, Cambridge, MA.

MARTIN WF (1974) The Managerial Environment, 1985, SM Thesis, MIT, Cambridge, MA.

MELENCIO PD (1974) An Analysis of the Planning System of Revere Sugar Refinery. SM Thesis, MIT, Cambridge, MA.

MICHALEK JJ (1974) Rate Setting in the Directly Placed REIT Commercial Paper Market. SM Thesis, MIT, Cambridge, MA.

MILLER SE (1974) Planning and Control in the Real Estate Investment Trust. SM Thesis, MIT, Cambridge, MA.

MONTIEL EL (1974) Financing the Growth of Development Banks in Latin America. SM Thesis, MIT, Cambridge, MA.

MORAN P (1974) The Industry Environment as a Determinant of Corporate Strategy Selection: An Empirical Evaluation of the Mobile Home Industry. SM Thesis, MIT, Cambridge, MA.

NORRMEN CJG (1974) Management Information Systems Design for Construction Operations. SM Thesis, MIT, Cambridge, MA.

PINTO JJ (1974) Valuation Problems in Mergers of Publicly Held Firms: The Premium Over Market Value. SM Thesis, MIT, Cambridge, MA.

POWELL DG (1974) Barriers to Innovation in the Construction of Buildings. SM Thesis, MIT, Cambridge, MA.

PUNJ D (1974) Opportunities and Strategies for Success in the Software Packages Industry. SM Thesis, MIT, Cambridge, MA.

RAMSELL CE (1974) Work in the Public Sector: Some Conceptual Notes on Job Satisfaction. SM Thesis, MIT, Cambridge, MA.

REID JC (1974) Computer Utilization in Higher Education. SM Thesis, MIT, Cambridge, MA.

RICE DW (1974) Situational Factors in the Success of Management Decision Support Systems. SM Thesis, MIT, Cambridge, MA.

RIGEL D (1974) POPS: Physicians' Office Processing System. SM Thesis, MIT, Cambridge, MA.

RING M (1974) An Analysis of the Portfolio Management Function in a Bank Trust Department. SM Thesis, MIT, Cambridge, MA.

ST. GERMAIN M (1974) Risk-Returns in the Stock Life Insurance Industry. SM Thesis, MIT, Cambridge, MA.

SAITO H (1974) Overseas Investment in the Australian Mining Industry. SM Thesis, MIT, Cambridge, MA.

SCHMIDT JK (1974) Strategy for Option Hedging. SM Thesis, MIT, Cambridge, MA.

SCHONBERG ML (1974) A Pension Fund Investment Study. SM Thesis, MIT, Cambridge, MA.

SHACHAT IJ (1974) A Parameterized Model for Selecting the Optimum File Organization in Multi-Attribute Retrieval Systems. SM Thesis, MIT, Cambridge, MA.

SNELLING HE (1974) The Utilization of a New Federal Health Manpower Program as a Catalystic Agent for Organizational Change. SM Thesis, MIT, Cambridge, MA.

SOLOMON HJ (1974) The Effect of Incorporation on Foreign Nationals into the Management of Multinational Corporations. SM Thesis, MIT, Cambridge, MA.

STANGLE BE (1974) An Analysis of Computer Security in an Ambulatory Medical Care Facility. SM Thesis, MIT, Cambridge, MA.

Vol. 7
1979

STRAKA RA (1974) Sulfur Pollution Taxation and Stack Gas Desulfurization Process. SM Thesis, MIT, Cambridge, MA.

SUSSMAN PN (1974) Supply and Production of Offshore Gas Under Alternative Leasing Policies. SM Thesis, MIT, Cambridge, MA.

SWARTZ DA (1974) France 1973: A Shift in the Macromanagerial Environment? SM Thesis, MIT, Cambridge, MA.

TAN JGS (1974) Currency Translation for Financial Control in the Multinational Enterprise. SM Thesis, MIT, Cambridge, MA.

TELSON ML (1974) The Economics of Electrical Power Generation Reliability. SM Thesis, MIT, Cambridge, MA.

VAIDYA R (1974) The Choice of Technology in Highway Construction Industry—A Case Study of Nepal. SM Thesis, MIT, Cambridge, MA.

VILLAMIL J (1974) A Cost–Benefit Analysis of the Performance of Development Finance Institutions. SM Thesis, MIT, Cambridge, MA.

WAGNER WR (1974) Strategies of Organization Development in the Securities Brokerage Industry. SM Thesis, MIT, Cambridge, MA.

WEIR IS (1974) Implementation Requirements for R&D Project Planning: A Case Study in the Pharmaceutical Industry. SM Thesis, MIT, Cambridge, MA.

WILLIAMS TV (1974) Perspectives on the Restaurant Industry, Restaurant Operation and Minority Restaurateurs. SM Thesis, MIT, Cambridge, MA.

WINTERHALTER DL (1974) Federal Government Managers in the Environment of Collective Bargaining. SM Thesis, MIT, Cambridge, MA.

WITZBURG JM (1974) The Anatomy of Defensive Medicine: A System Dynamics Analysis. SM Thesis, MIT, Cambridge, MA.

WOLFE JE (1974) Business Opportunities in the Videocassette Industry. SM Thesis, MIT, Cambridge, MA.

WOLTER JW (1974) Scale Economics in the Mining Industry. SM Thesis, MIT, Cambridge, MA.

WYMAN J (1974) Perceptions of Career Opportunities by Male and Female Managers. SM Thesis, MIT, Cambridge, MA.

YANO T (with HARDY D) (1974) The Acquisition as a Form of Foreign Direct Investment in the USA: A Study of the Manufacturing and Banking Industries. SM Thesis, MIT, Cambridge, MA.

YONKER RC (1974) Plant Influence on Capital Allocation in a Divisional Environment. SM Thesis, MIT, Cambridge, MA.

YOSHITOMI M (1974) The Internalization of Japanese Firms Concentrating on Organization and Labor Strategy. SM Thesis, MIT, Cambridge, MA.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 2. pp. 89 to 99
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79 0401-0089$02.00/0

EDITORIAL

Ackoff's Fables

FABLES AND MORALS

NOTHING is calculated to impress students and observers of the managerial scene more than a simple story with a moral attached to it. You can expound on principles of model building until you are blue in the face, you can philosophize about systematized procedures of investigation, you can contrive some scintillating theories about the normative way of life of managers and management scientists; but tell your audience the story of a particular episode, of how as an analyst you came to grapple with a problem set by 'the authorities' (your boss, or your client), how you managed ingeniously to reformulate the problem and the objectives of the study, how you unravelled a solution—so obvious and yet so elusive—and how you triumphed against all odds to have your prescription accepted, tell them all that and the chances are that the lessons to be learnt from this experience will be imprinted on the memory of your captivated listeners more firmly than all the generalized abstractions that you can muster.

Story telling as a means of conveying information or involving the active participation of students is, of course, not new. Harvard and other schools of business administration (not to mention schools of law and medicine) have thrived on the case method for decades, and it remains a most successful recipe for passing on experience to the uninitiated. Furthermore, story telling is both a convenient and a flexible mode of communication. A story can contain as much detail as is thought appropriate to make a full impact, the details and the record of events can be suitably modified, the story need not necessarily relate your own experience, nor need it be true (at least not in every respect). Since its prime objective is to convey a 'message', you feel that you are justified in taking licence in constructing the story to suit your purpose, provided, of course, that you do so from the best of motives and that you do not deliberately misrepresent the facts. We all indulge in this game—teachers, consultants, managers—some of us play it more expertly than others, some are more brazen and much more effective.

Russell Ackoff is undoubtedly one of the great masters of this art. Whether he addresses gatherings of enthusiasts or semi-sceptics, or whether he writes to acquaint a wider audience with the mysteries and promise of management science (or "systems science" as he prefers to call it), he is conscious of the need to cite examples from a rich repertoire of case histories. Each anecdote is expertly constructed with assiduous economy to highlight, with the bare minimum of detail, the main ingredients of an intriguing problem and the compelling arguments for a particular method of approach or preferred solution.

A collection of these anecdotes was well overdue, but now Ackoff has obligingly assembled them in his recent book "The art of problem solving", appropriately sub-titled "Accompanied by Ackoff's fables" [1]. In fact, the anecdotes so dominate the narrative, that the titles should have been reversed to read "Ackoff's fables, accompanied by a commentary on problem solving", and (for the reasons that I have already referred to) the stories constitute the more fascinating part of the book. None of the case studies is new; Ackoff has told and re-told them on numerous occasions, and although he can vouch for most of them, he readily admits to some being hearsay and possibly inaccurate. But as one of his friends once said, "If they are not true, they ought to be" [1, p. 17], and this is perhaps the most apt description of their role and impact.

Vol. 7
1979

The author does not, of course, recount his fables and many experiences just for the fun of it; he has a much more serious purpose in mind, and although his commentary on the art of problem solving is in contrast somewhat turgid and occasionally rather dogmatic, his attempt to generalize from the particular is both relevant and commendable. If the case study approach is a legitimate means of transmitting experience in an orderly fashion, what lessons can we learn from vivid yet often incomplete accounts of past episodes?

In the field of medicine case histories are chosen with a view to providing as comprehensive a classification as possible of various diseases and accounts of past treatments, so that when a medical practitioner is faced with a problem he can search for an identical previous case to determine what remedy to prescribe. Similarly, in the field of law, where past court decisions and interpretations are of prime importance, case law is indispensable to the lawyer when he is asked to predict judgements and advise his clients accordingly. Of course, both in medicine and in the law new cases arise which do not exactly replicate the past, but even then (and depending on how novel the new situations are) doctors and lawyers are primarily guided by their knowledge of the past.

Can the same be said about business problems? Those who believe in the omnipotence of the study method firmly assert that it is the only way in which an adequate appreciation of the business environ-

ment can be achieved and in which the possible outcomes of alternative decisions can be assessed. Even those who argue that the managerial scene is too dynamic and governed by too many variables to be neatly categorised by a finite number of case studies, would readily agree that theory without experience, be it even secondhand or simulated experience, is both sterile and dangerous, and hence would support the use of case studies to illustrate the application of various principles or methods of investigation. And so, whether you are an avowed supporter or a sceptical opponent of the wide use of case studies, you are bound to dwell on the lessons that may or may not be drawn from them.

Over the years, managers and analysts tend to generalize from their experience and develop their own pet theories about the behaviour of the business environment and its likely responses to certain managerial actions. With the accumulation of experience these theories are postulated with ever increasing conviction, nurturing (and in turn being reinforced by) various prejudices and attitudes. It is only natural, therefore, that when you read about the experiences of others you tend to accept or reject the conclusions proffered, depending on whether they conform to your theories or not. Thus, readers will react differently to Ackoff's fables, they will find succour in some of his stories and be rather sceptical about others. In this respect I am no exception, and my own assessment of Ackoff's fables reflects a subjective judgement that others, and at times even Ackoff himself, would disagree with, but it is by comparing experiences and arguing about our deductions that we are likely to gain a better understanding of our environment, and perhaps even rid ourselves of our prejudices.

OBJECTIVES AND CONSTRAINTS

The first noteworthy lesson that I would commend to others is that *objectives are not what they seem to be.* More often than not, a stated objective reflects a narrow framework with a limited number of decision variables, hemmed in by a host of constraints. The more the analyst delves into a problem, the more he realizes that the stated objective would involve an exercise in sub-optimization, or that it would be incompatible with the goals of other related sub-systems, or that it is an incoherent expression of some hidden and perhaps more sophisticated objectives.

And herein lies one of the dilemmas of the analyst. He cannot proceed with the task of finding a solution to a problem before he knows what the problem is and what the solution is supposed to achieve, and at the same time he is unable to ascertain what the objective is (or should be) without tackling the problem and prodding the principal characters on the scene. This confusing state of affairs is well exemplified by Ackoff's very first fable, concerning the small company

producing hand tools for carpenters and seeking to diversify its operations. The consultant (whom Ackoff identified as Aesop, no less) came up with several proposals, all of which were rejected by the president of the company, who was looking for something better. "Aesop asked what 'better' meant. The president said that he could not possibly define it but that when Aesop came up with an alternative that was 'better' the executives would recognize it easily" [1, p. 22]. At least in this case the client freely admitted that he was unable to specify the criteria by which solutions were to be assessed, but all too often such an admission is not forthcoming, or the client initially feels that he knows what he wants, so that the problem has to undergo several transformations before the analyst fully understands the purpose of the model building exercise.

Ackoff is at pains to explain that ends (objectives) and means (courses of action in pursuit of ends) are relative concepts [1, p. 20]. A particular end can always be regarded as a means to achieve another end. Any analyst who has had the experience of probing executives in search of their motives knows that by repeatedly posing the question 'why?' he embarks on the road of what first appears to be greater clarification and more precise quantification, but then becomes fog-bound by general and unquantifiable aspirations, such as 'happiness' and 'self-fulfilment'. Consider, for example, the following dialogue (I have illustrated this point elsewhere [3]):

- Mr Executive, why do you want me to examine the effectiveness of your advertising budget?
- Because I want to ensure that it is cost-effective.
- Why?
- Because I want to increase the profit of this company.
- Why do you want to increase profit?
- If profit increases, the board will invest more, the company will expand and grab a dominant share of the market.
- But why do you want to achieve that?
- Because it will secure our future, make us less vulnerable, give greater career opportunities to my managers.
- And what is the purpose of that?
- We shall have a sense of achievement, the satisfaction of winning against odds, *etc.*

Why, why, why? The analyst soon realizes that further probing will lead him nowhere, and knowing when to stop this quest is an art that cannot be formalized in a textbook. Furthermore, as the process of clarification continues, he begins to have doubts about the meaning and validity of some of the answers: If, as stated in our example, the executive wants a secure future, does it follow that the company needs to expand its market to achieve that aim? And if the question

Vol. 7
1979

'why?' elicits several ends, what do you do when one end is neither the result of nor the means to another, or when they are patently incompatible? Thus, the series of responses in this dialogue between investigator and executive is often inadequate as a guide in the ensuing analysis, for the simple reason that a single objective is rarely sufficient to describe the system's aspirations. And in the same way that for each end it is generally possible to identify several alternative means, each means leads to several end results with their attendant repercussions.

If the analyst is left wondering how to distinguish between ends and means, he is certainly made aware of the close relationship between objectives and constraints. Ackoff repeatedly makes the point that many of the constraints are self-imposed and that a searching reappraisal by the analyst often reveals an astonishing degree to which constraints can be reformulated, relaxed, or disregarded altogether:

> "Many of the constraints that limit what we can do in a problem situation are unconsciously self-imposed. By raising them to consciousness and questioning their validity, we can often find ways of removing or relaxing them. This can sometimes be done by finding new ways of using available technology, by developing new technology, or by using our indigenous technology that is overlooked because it is in a place we consider unlikely, for example, among the uneducated" [1, p. 77].

Indeed, it has often been said that more substantial payoffs can generally be achieved by attacking constraints than by concentrating on the manipulation of controllable variables for the purpose of improving the value of some objective function. Furthermore, as I have said elsewhere, "the distinction between goals and constraints is an artificial one. That this is the case in a satisficing problem is obvious enough, but even in a problem of optimizing the constraints may be regarded as an expression of management's desire to have minimum attainments of performance with respect to various criteria. All (internal) constraints are, therefore, expressions of goals" [4, p. 185].

The question arises, why are they then expressed in the form of constraints rather than explicitly as objectives? There are two answers to that: First, the specification of constraints has the implication of a hierarchy of purposes, so that certain modes of operation and performance measures in a satisficing format are assured whatever further improvement in others can be obtained; and when the analyst begins to question the validity or level of stringency with which the constraints are imposed, he finds that managers are more amenable to the relaxation of some than others, thereby strengthening the evidence that a hierarchy of constraints often exists. Secondly, internal constraints are imposed as a means of controlling the area of discretion of subordinates or sub-systems [4, chapter 9], or as a means of limiting the level of intervention from peers and other functions of management [4, chapter 8]. These may be identified as vertical and horizontal constraints respectively, and they have one feature in common, namely

a desire by the executive responsible for generating or formulating them to maintain or to expand his room for manoeuvre and freedom of choice (there may, of course, be a group of executives or a whole section of the organization structure with a vested interest in this respect).

Thus, the removal or relaxation of constraints is not as easy a task as it sounds. Many naive analysts have fallen into the trap of believing that a rational and dispassionate discussion with executives about constraints would soon reveal their superfluity or lack of consistency, and that as a result the executives would readily agree to drastic modifications. What the analysts fail to recognize is that behind what they perceive as a questionable rationale there are motives of self-preservation and ambition, and that unless these can be accounted for and suitably satisfied, resistance to change (even hostility to the efforts of the analysts) is bound to ensue.

Vol. 7
1979

CAUSE AND EFFECT

In an attempt to understand the behaviour of a system, the analyst has to explore the relationship between variables, either for predictive or for manipulative purposes. If he does not know how to predict the value of variable *A*, he searches for another variable *B* which is strongly associated with *A* and which is easier to determine, so that once *B* is known the future value of *A* can be inferred. A good example is the desire of financial institutions to predict interest rates or currency exchange rates, and they are constantly on the lookout for possible associated predictions on which they can base their inferences. As for his manipulative purposes, the analyst wishes to know what would cause a change in the value of a particular variable, so that he can contemplate taking action to bring about such a change. This is where Ackoff offers some valuable advice:

> "Our ability to solve problems depends critically on how well we conceptualize the causal connections between what we do and what we want. Many of our problem-solving failures derive from assuming a causal connection where it does not exist or incorrectly characterizing a causal connection where it does exist. Perhaps the most common single cause of failure in problem-solving derives from incorrectly assuming a causal relationship between variables that have only been demonstrated to be *associated*. Variables that tend to change together, in the same or opposite directions, are associated." [1. pp. 101–2].

I am reminded of the story (alas, I cannot recollect its source) of the boy who is asked why trees grow. He recalls that in the spring birds sit on tree branches and from time to time burst into song, and so he concludes that the singing of the birds is the reason why trees grow.

Ackoff then recounts two examples. The first is the attempt, often made, to relate a company's total sales to its advertising budget. When the one is plotted against the other over a number of years, a high statistical positive correlation is often found, clearly showing a strong association between the two variables. But Ackoff warns us that it does not follow that there is a causal relationship here and that the company sales will necessarily increase if more money is spent on advertising. In one case he plotted the annual sales against advertising expenditure in the *following* year and obtained an even stronger association, but it would have been difficult to argue a causal relationship where the effect clearly preceded the alleged cause. In fact, the strong association in that case was simply a reflection of the procedure adopted by many companies of fixing the advertising budget as a proportion of last year's sales, or as a proportion of the expected current year's sales (which in turn is related to last year's). So that, if anything, the inferred causal relationship should be reversed, with the advertising budget identified as the effect and not as the cause.

The second example is one which has been the centre of a raging controversy for many years. It concerns numerous studies that have been carried out to determine a causal relationship between smoking and lung cancer, and Ackoff states categorically that "smoking may be a cause of lung cancer, but their correlation is not an adequate basis for asserting that it is" [1, p. 103], a sentiment that many readers will take issue with. To illustrate this point, his problem solver, Aesop, showed how data for *per capita* of consumption of tobacco, used in a paper published in a prominent journal to infer a causal relationship with the incidence of lung cancer in several countries, could be similarly employed to reveal a negative correlation with the incidence of cholera, so that if inference on the basis of these studies was at all appropriate, he was right in suggesting that smoking prevents cholera. Now, he readily admits that such a conclusion would be facetious, but argues that the postulated causal relationship between smoking and lung cancer is equally open to criticism.

For Ackoff, an association by itself can never prove causality, but the analyst is bound to ask whether his confidence in a particular proposition should not be enhanced when he is presented with the results of several alternative studies, coupled with a plausible explanation for the causal relationship. After all, certain schools of the philosophy of science argue that no proof of any theory ever exists, and that we can justify our belief in its validity only as long as it has not been refuted by observation and experiment. The arguments for and against the theory that smoking causes lung cancer thus clearly demonstrate that what seems an open and shut case to some is less than convincing to others.

And yet, decision makers need guidance for action. The search for causal relationships is not a mere academic exercise; it is a prelude to contemplating whether decisions should be taken with the expectation that they will bring about a desirable change. In the case of smoking, for example, the Government is faced with several alternatives, from doing absolutely nothing to banning the manufacture and importation of tobacco-based products altogether. Whatever action it decides to take, it must be based on some rationale regarding the detrimental effects of smoking, and there is always the possibility of the two classical types of error occurring, type 1 relating to expensive action being taken but having no positive results and type 2 of no action taken while the ill-effects, which could have been avoided, continue to persist [1, p. 47]. In the end, it is a matter of judgement as to whether an established association between variables should be acted upon or ignored; it would certainly be difficult to justify the maxim that no action must ever be taken before proof of causality has been fully established.

Vol. 7
1979

PROBLEMS AND SOLUTIONS

When is a problem a problem, and when is it a solution to another problem? I was struck by Ackoff's study of alcoholism as an example of how often we tend to prescribe remedies designed to eradicate symptoms but not the cause. Alcoholism, says Ackoff, "is usually treated as a *problem* to be removed rather than as the alcoholic's *solution* to a serious personal problem. No wonder that when his 'solution' is denied him he often returns to it as soon as he can or develops other 'solutions' that are equally damaging. The alcoholic's original problem is usually left unresolved when he is denied alcohol or is induced to abstain" [1, p. 116].

Thus, any state of affairs which we are called upon to remedy may be viewed as a means that has evolved (or has been specifically adopted) in order to achieve certain ends and thereby solve some prior problems. And similarly, any solution developed by an analyst will abate certain complaints and create new ones. It is foolish for the analyst to imagine that once a cure is prescribed and the pains go away, the story is then complete with a comforting happy end. Just as problems cry out for solutions, so solutions beget new problems. Some of the consequences of proposed solutions can be predicted and the system can be duly prepared to handle them, but some cannot; they develop as the environment changes and adjusts to the new values of the controllable variables, so that while the performance of the system may initially come up to expectation in certain respects, new strains emerge that require attention. These unpredictable consequences are the analyst's nightmare; no wonder he has to pause from time to time and ponder whether the disease is not preferable to the proposed cure.

Experience and case law are obviously helpful when the analyst has to make this judgement, but drawing conclusions from other situations can be a risky business, and in this respect Ackoff is sometimes too hasty in jumping to conclusions. For example, in his discussion of the drug addiction problem in the US he contrasts the extreme measures taken by his Government with the experience in Britain, and says: "By legalizing the use of addictive drugs and dispensing them in a controlled way, England has had much more success than America in treating addiction, and it has avoided the huge criminal problem created by the demand for drugs that are illegal" [1, p. 117]. Many will argue that this is a gross oversimplification of the success or otherwise of drug control in Britain, and it is certainly open to question whether legalizing drugs in America will solve its drug problem.

Ackoff urges analysts to be imaginative, and it is obvious that creativity in problem solving is likely to have a more significant impact on a system than some minor adjustments in controllable variables within a narrowly defined framework. The dilemma for the analyst is that the bolder his solution, the more difficult it is to predict the results, not only because of lack of evidence as to how the system will behave under new operating conditions, but because the whole internal environment may be so drastically affected that many of the assumed functional relationships between variables no longer hold.

Of course, from time to time a seemingly outrageous solution is proposed, only for the analyst then to discover that it has already been tried or adopted elsewhere. Ackoff tells the story of a group looking into the problems created by the rapid increase in the number of cheques that need to be processed and cleared by the US banking system. From suggestions as to how cheques could be eliminated altogether by the electronic transfer of funds between bank accounts, the group came to realize that if a bank account includes a record of all transactions of income and expenditure, it could be used to prepare income tax returns; to avoid the difficulty of individuals having more than one bank account, the group then came up with the ingenious idea of a new tax system based on expenditure rather than on income [1, p. 28]. The idea though, is not new, and apart from Sweden and England which Ackoff mentions as countries where "such efforts were being made", value added tax is wide-spread in the Common Market and many other countries, and sales tax is well established in many states in the US. The total abolition of income tax and its replacement by a tax on consumption is a political (not an administrative) problem, and its pros and cons have been hotly debated by political parties in Britain for some time now. What is ingenious and exciting to some is commonplace to others.

THE POWER OF INCENTIVES

Of the many fables recounted by Ackoff I particularly like two which demonstrate how clever yet simple incentive schemes can be designed to influence behaviour and thereby bring about a change in a desirable direction. Analysts tend to devote their attention to a limited number of prescribed controllable variables and do not realize that some of the principal characters that they have to deal with can be converted from being overt opponents to a proposed solution to becoming its avid supporters.

One of his stories concerns the amount of spare parts carried in vans by servicemen working for an operating company of the Bell Telephone System in the US [1, p. 59]. Since each repairman was paid according to the number of completed calls he made on subscribers, he filled his van with parts to meet any conceivable eventuality, with the result that very excessive stocks were carried about in these vans.

A conventional inventory control exercise would have concentrated on working out the average usage of parts and the probability of runouts, so that specified stock levels could then be enforced on the servicemen. Instead, a new scheme was devised for payment to the repairman to increase with the number of calls made, but to decrease with the value of stocks carried in the van. In this way, he had a positive incentive to reduce the stock levels, and he did so voluntarily without the inevitable hostility that he would have felt towards an imposed prescribed solution.

The second example relates to the commonly found case of a company producing too great a variety of products, leading to intractable scheduling problems on limited manufacturing facilities and to low profitability [1, p. 69]. The solution—short of adding production capacity—is to reduce the variety of products, and as one often finds that a relatively small number of products is responsible for the major part of turnover and profit, it is not difficult to identify non-profitable products that should be dropped from the range. In the example described by Ackoff, the arguments that one usually encounters in such cases were that the customers for the products in question were too important to ignore and that removal of the low volume products would adversely affect the sale of those that were highly profitable.

Instead of attempting to refute these allegations, a new payment scheme for salesmen was set up to replace the previously flat commission on total sales with one that had a higher commission for profitable products but none for unprofitable ones. The salesmen devoted their efforts to selling accordingly, their earnings increased substantially, so did the profit for the company, and many unprofitable lines were automatically eliminated.

Vol. 7
1979

BETTER THAN WATT?

I could go on, but I won't. There is no single moral to be deduced from Ackoff's reminiscences, except that problem solving is not an occupation that can be prescribed by some mechanistic well-trodden procedures. Even the terms that we use to describe what we do are open to several possible interpretations, and it is not always immediately clear, particularly at the strategic level, what the problems really are, let alone whether they have been solved.

Perhaps that is why 'story-telling' is so alluring, and yet so deficient. It encapsulates the frustrations and triumphs of the story-teller in a most captivating way, but the inevitable process of simplification that has taken place in order to isolate the grain from the chaff leaves many questions unanswered, particularly questions about alternative approaches and the effectiveness of other possible solutions. Furthermore, I am often curious about what happened after the end of the story: If it ended happily, did the main characters continue to live happily ever after, and what related new problems did they have to face? And if there was no happy ending (Ackoff recounts many cases where his recommendations were rejected), I would like to have been able to speculate on how methods could have been devised to change the course of events. If we pride ourselves on being problem-solvers and we find that our clients are stupid or obdurate and will not accept our advice, that too is a problem, and it needs to be tackled.

In any event fables like Ackoff's do demonstrate that managerial problems need attention and that we have a contribution to make. I hope that Watt is listening, and if he is not, he should be.

"But if Watt's mouth was open, and his jaw sunk, and his eyes glassy, and his head sunk, and his knees bent, and his back bent, his mind was busy, busy wondering which was best, to shut the door, from which he felt the draught, on the nape, of his neck, and set down his bags, and sit down, or to shut the door, and set down his bags, without sitting down, or to shut the door, and sit down, without setting down his bags, or to set down his bags, and sit down, without shutting the door, or to shut the door, from which he felt the blast, on the nape, of his neck, without setting down his bags, or sitting down, or to set down his bags, without bothering to shut the door, or sit down, or to sit down, without troubling to set down his bags, or shut the door, or to leave things as they were, the bags pulling at his hands, the floor pushing at his feet, and the air puffing, through the door, on the nape, of his neck. And the conclusion of Watt's reflexions was this, that if one of these things was worth doing, all were worth doing, but that none was worth doing, no, not one, but that all were unadvisable, without exception." [2].

SAMUEL EILON
Chief Editor

REFERENCES

1. ACKOFF RL (1978). *The Art of Problem Solving, Accompanied by Ackoff's Fables*. John Wiley, New York.
2. BECKETT S (1953). *Watt*. Olympia Press, Paris (The extract comes from the 1976 edition, published by John Calder, London, pp 219–220).
3. EILON S (1978). The law of the final cause (Editorial). *Omega*, **6**(1), 1–3.
4. EILON S (1979). *Management Control*, 2nd edition. Pergamon Press, Oxford.

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 2 pp. 101–103
© Pergamon Press Ltd 1979 Printed in Great Britain
0306-0483/79/0401-0101$02.00/0

FEEDBACK

People and Size

I AM GRATEFUL to the Editor for the chance to reply to his challenging article: 'A Question of Size' (Vol 7 No 1)

It was Dr Schlesinger who said that one of his children sat for an examination in which one question read: 'Discuss the causes and consequences of World War II (use the other side of the paper if necessary)'. That is how I feel when trying to cope with this vast subject in a limited space. Both Professor Eilon's article and my paper, to which he referred (Problems of Size and Efficiency in Industry Today [*Omega* 7 (1), 1–7]), majored on the problem without getting to grips with the solutions. At least, not in detail.

Let me point towards some of those solutions.

I make no apologies for size in my industry. Without adequate economies of scale the manufacturers of most types of motor vehicle will not survive no matter how exciting their products or generous their benefactors. We can give up on size in this country if we are prepared to make substantial reductions in our standard of living.

However, the purpose of this note is not to defend size itself but to look at how to solve some of the problems that size creates.

Schumacher is basically right, and, even in industries like mine where size is inevitable, much of his theory can be applied. Even with our Europe-wide economies of scale and homogeneity of manufacturing processes it is possible to see much of his five principles at work in my own company. To go over them briefly:

(1) *Subsidiary function.* Of course, management cannot be given freedom in the individual plants to make anything they like, and strict targets and standards have to be set in each plant to produce the parts for particular vehicles or components. Nevertheless, each manufacturing facility identifies strongly with its particular product and has many of the characteristics of an independent smaller supplier. Locations like the Southampton truck plant, Thames Foundry or the Swansea axle plant can certainly not be defined as suffering from overcentralisation or lack of individual character.

(2) *Vindication.* Each plant has its own given criteria, which includes its own schedules and budget.

(3) *Identification.* In most plants employees identify more strongly with their plant than with Ford—certainly in the UK. If you have any doubts, try walking on the shop floor at Halewood and telling them they are all Europeans! While every effort is made to explain their reliance on each other, the individuality of each plant can be a benefit to the company and is a potential to be harnessed. Indeed, identity is encouraged by areas of management within a single location.

(4) *Motivation.* This is the biggest problem and I will deal with it in more depth later in this note.

(5) *Middle axiom.* Depending on what Schumacher sees as the right balance between order and entrepreneurial activity, we don't think we do badly. Successfully designing, making and selling motor vehicles requires a powerful mixture of flair and rigid discipline. Read Edouard Seidler's book 'Let's call it Fiesta', and see what you think.

Clearly, a company like mine can only go some of the way towards satisfying the Schumacher criteria for health and happiness. His principles would be much easier to apply to a large conglomerate, say, than a highly concentrated industry like motor manufacture.

Given these constraints we devote untold time and effort to coping with the problems of size, especially where they affect the individual. The popular accusation that we 'don't care' cuts deep, for it is both manifestly unfair and superficial. If you dare to say that the firm application of a system is needed—or even to admit that some jobs in mass production are unavoidably less interesting than others—you are branded as a slave-driving revisionist. Experience also shows that an uninteresting job to one man can be the exact opposite to another. People's interests and aspirations differ more than the jobs themselves.

But if you are to survive with our size, and in our industry, the first thing you must do is face the facts.

There is an old story of the beetle meeting the centipede and asking how he managed to walk with so many legs. "Why, it's easy", said the centipede, thinking about it for the first time, "I start with this one and . . ." The next day the beetle came back and found his friend the centipede dead from apoplexy.

The moral of that story is that every one of your legs is vital, no matter how tiny, but even more vital is that they work with a common purpose. It means that every member of the team—be he management, union or individual—must sacrifice some aspirations and work in a commonly agreed direction. With the usual (and totally incorrect) cries of 'union basher!' ringing around me, it must be pointed out that the system works extremely well in every country of Ford's worldwide operations, but that in Britain life is complicated by a unique state of affairs whereby the trade unions have considerable power over their employers but considerably less over their members.

All corporate forms of existence, from the ant colony to the multinational company, must have a central discipline if it is to survive. The bigger the organisation the more rigid and detailed the system becomes and a degree of bureaucracy comes into virtually every aspect of a big international manufacturing operation. In this respect the biggest blessing is free enterprise. As long as you have choosy customers who are cost conscious, and a few strong competitors, even the mastodons can maintain a relatively lean and hungry look.

Given this sort of background most of the problems of size can be overcome so long as there is a firm sense of purpose and firm pursuit by management of the company's objectives and intelligent use of procedures.

The major problem area that is left, however, is how within that system to foster the vital motivation and identity of the individual. The enlightened industrialist has realised for some years that the customer and the individual employee are his two most vital commodities. Many companies have been caring increasingly for these two constituents since long before the recent flush of intellectual interest and white papers.

If we are guilty of anything it is that we haven't yet got it all right, not that we aren't trying. At least, we believe that, thanks to exhaustive research and design, our products satisfy the customer more than most, but employee relations still have a long way to go.

We have also *come* a long way, though. Years of experience and constant improvement have helped us identify six areas of human involvement on which modern management must concentrate.

The first, and probably the most difficult, is to try to make the jobs themselves more interesting and—as my American colleagues would say—more meaningful. When manufacturing on a really big scale you cannot escape from the production line, but you can replace many of the more dull and repetitious tasks with automation. This is still far from perfect, but at least we recognise the fact. We are as unhappy about the existence of less attractive jobs as some of the people themselves who actually do them.

A degree of amelioration can be achieved by some movement of jobs, both sideways and upwards. An employee doesn't *have* to spend his life doing the same task at the same spot. Within a limited context people can move from one job to another, and manual employees are encouraged to qualify initially as training reserves for more demanding jobs, thus facilitating their eventual promotion to those jobs.

Incidentally, we rapidly learned that not everyone *wants* to move. Do not forget that there are some people who for a variety of reasons *prefer* doing a repetitious job, and forcing them to move around is as inexcusable as neglecting them and is considered by them as cavalier treatment.

Thirdly, there is the establishment of a caring management who is aware of the human factors. This is made easier when the members of management have themselves had to work their way up through the various disciplines (and have good memories). It is important for plant and central management to be backed by a dedicated industrial relations team. It is also important to pay attention to health and safety over and above statutory requirements.

We then come to the highly fashionable world of communication. Far from being the panacea it is held to be, good communication is just one crucial aspect of good management. For several years we have been devoted to a very comprehensive open communication programme with all employees, and we know it helps.

The annual figures and several other aspects of the company are presented and discussed with all employees on a regular basis. We have a fortnightly company newspaper, regular newsletters, Joint Works Committees, departmental meetings, suggestion schemes and a host of other methods by which everyone is kept fully in the picture on everything from production targets to the section darts results. Training and education also play an important part. More could be done by industry to improve the technical and general education standards of its people. This can be as refined as running 'high flier' courses for promising young engineers to something as basic and essential as language classes —in the company's time—for employees with language difficulties.

While taking care to educate rather than indoctrinate, it is important for everyone working in industry to learn some of the basic facts of industrial life, such as the need for profitability and where profits are allocated. How can you work with any conviction if you continue to believe that all the profits go into the directors' pockets?

And, lastly, it is essential that people are adequately rewarded for their efforts, not just financially but also with decent pensions, catering and social facilities, a high degree of job security, and recognition. In an age of increasing greyness, where the social and political ethic is one of everyone getting the same regardless of their contribution, this creates more than its share of headaches. But at the end of the day the pay packet must reflect the amount contributed by the individual.

Those, then, are the six key areas of human involvement to which management efforts are increasingly directed: making jobs more interesting; job movement, both around and upwards; a caring and aware management; effective communication; education; adequate reward.

While size creates problems we do not hear enough nowadays of its good points. In addition to commercial viability, one frequently overlooked aspect of size is that many, if not most, people actually prefer to be part of a system despite the disciplines and frustrations involved. People also like to be part of a successful operation. A man from Mars would probably find the Army, with its plethora of regulations, quite ridiculous. I know from my own military service that the system sometimes drives you mad. Yet I hear nothing but fondness in the voices of those of all ages who have been subjected to that monolithic system. We are largely creatures of habit who like to know where we stand in the order of things. We are gregarious and like security. It is not always a bad thing when some of it rubs off on industry.

In conclusion, I cannot claim to have all the answers on this difficult subject, and it's easy to criticise what has been done. As practical people in my Company we are trying. The Schumacher points, simply repeated by Professor Eilon, are easy to make, difficult to solve.

We are now living through a time of rapid change, and aspirations for work improvement do seem at times to be running ahead of our ability to satisfy people's material ambitions. Small would be beautiful if it also applied to income and the real standards of living. This is most emphatically not the case in Britain today. The explosion of dissatisfaction we have seen in the past six months in both public and private sectors of industry stems from the unacceptability of our present level of wealthy creation which has not grown for four years. These are perhaps more serious problems than those of size at this stage of our technological development and compromises have to be made today, perhaps, more in favour of effectiveness than those described so well by Dr Schumacher and Professor Eilon.

SIR TERENCE BECKETT
The Chairman and Managing Director
Ford Motor Company Limited
Eagle Way
Brentwood
Essex CM 13 3BW
UK

Vol. 7
1979

Information as a Tangible Benefit

THREE contributions in a single recent issue of *Omega* [1–3] raised the issue of the benefits of information to managers. They all in various ways dodged perhaps the only major outstanding issue in MIS Management Information System. No amount of Mathuszian finesse [3] in measuring the incremental productivity of MIS projects as a basis for choosing between them answers the questions: 'What good do any of them do, and to whom?'.

The usefulness and benefits of information can only be realised by individual managers in particular situations. Any attempt to devise arbitrary measures of benefit for information services that does not start from and end with an analysis of the situations in which individual managers find themselves must be poppycock. For this reason the survey by Higgins and Finn [1] is encouraging. They are seeking to find out the attitudes which individuals (in their case Chief Executives) have towards corporate models and information. Mace [2] writing on accounting addresses the major issue but simply cites Snavely's [4] listing of attributes which data must have to be useful or beneficial. He is trying to demolish the case that such reports can be optimally designed if they have predictive ability.

Whilst it's all very fascinating—academically, so to speak—why on earth do we not get down to the situation specific study of the ways individual manager's needs and/or wants for information arise and the outcomes that possession thereof occasions. Marketing and social/psychological researchers spend a very great time indeed analysing such situations. Good marketing research *always* proceeds through a qualitative stage before attempting to generalise and so of course, do all effective MIS creators and operators.

The first discovery we always make is that each individual approaches a similar decision making situation with different expectations, attitudes, risk preferences and the like. That initial approach stance is the outcome of past experiences and learnings as well as the perceived pressures of the current position and the situation looking forward. Taking that stance apart, recently, for four individuals heading up identical management teams, revealed an infinite variety of needs responsiveness and benefit opportunities. And one individual found he could only absorb the information well in picture form (a sociologist); another thrived on numbers (a willing engineer). The medium, if not the message, became a crucially important element determining whether any benefit could result, what could be termed the 'massage'.

The arena in which the individual manager makes use of the information he has assimilated also has to be studied in detail. His behaviour must be studied—preferably in comparison with those who do not have similar information but, if this cannot be accomplished, the individual's attitudes and behaviour must be studied longitudinally. It's a daunting prospect for any who are used to working with inanimate machines or who like to believe that *their* value judgements about what is good for a manager are best. The marketing fraternity call that production orientation and find it works with customers if you are lucky or lucky enough to be empathetic.

MIS cost streams in a business, thanks to more than a century of accountancy work and despite some theological divergences of opinion about allocation procedures, *are* relatively speaking easy to grapple with. The changes in managerial workloads, the shift of time allocation to greater or lesser opportunities, the posing of tactical and/or strategic questions that had hitherto gone unconsidered, these can only be assessed if we believe we should try. We find that the arrival of information frequently creates a need for more information—either for greater benefit to be achieved or because the first information sets undermined previously held confidence. Whether any MIS can be treated on the basis of five-year life-cycle projections that Mathusz suggests is tendentious. Our own experiences are that two or three would be more realistic with costs following the old bathtub curve over that span. Rather like the advertising budget, MIS is probably best handled on a dynamic budgeting approach from a zero base each time.

To conclude, Higgins and Finn [1] have started the ball rolling by using a marketing research approach. Good. Now let's go further and use the techniques developed in the 60s and 70s. Quintessentially, let's study the Information Consumer with all the skill that the marketing researcher studies the industrial purchasing officer or the housewife. The technique exists. Aux armes.

P.S. One of us has started already with university libraries [5]. Since 1972 we've been researching the question: "How much 'good' does a library do?" It's the opposite view from "How good is the library?" And both of us have faced the challenge in our own international publishing house with our own MIS.

JUDITH ATHERTON

Management Centre
University of Bradford
Bradford, W. Yorkshire
England BD9 4JL

GORDON WILLS

School of Management
Cranfield, Bedfordshire
England MK43 0AL

REFERENCES

1. HIGGINS JC & FINN R (1977) The chief executive and his information system. *Omega* **5**(5), 557–566.
2. MACE JR (1977) Criteria for the selection of reporting method. *Omega* **5**(5), 567–582.
3. MATHUSZ DV (1977) The value of information concept applied to data systems. *Omega* **5**(5), 593–604.
4. SNAVELY JH (1967) Accounting information criteria. *Acc. Rev.* **42**, 223–232.
5. OLDMAN C & WILLS G (1977) *The Beneficial Library*. MCB Publications, Bradford, Engl.

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 2. pp. 105 to 112
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0401-0105$02.00/0

Technical Change and Employment in the Post-War Gas Industry

RUSSELL MOSELEY

University of Aston, UK

(*Received August* 1978; *in revised form October* 1978)

The development of the post-war British gas industry reveals a pattern of continuous productivity increase accomplished by a series of significant technological changes and substantial reductions in the labour force. Moreover, the transition from coal gas manufacture to the gassification of oil, and, more recently, to the exploitation of natural gas was accompanied by marked changes in the composition of the work force. Such changes raise a number of questions concerning the potential for re-training and redeployment measures within industries undergoing technical change. The paper examines the manner in which the shedding of labour took place within the gas industry, looks at the extent to which retraining measures were a feature of its development, and suggests some more general issues that are raised by this case study and which warrant further consideration.

INTRODUCTION

WITH THE imminent arrival of spectacular advances in microprocessor technology, the relationship between technical change and employment has once again become a topic of major concern. Lip service may be paid to calls for the adoption of labour saving technologies in an attempt to increase productivity, but a figure of one and a half million unemployed in Britain has ensured that neither labour nor management can be unaware of the broader implications of such technical change. The central dilemma remains: if the country is to remain a competitor in international trade, can the opportunities presented by new technologies be ignored? On the other hand, will new economic opportunities be sufficient to avoid resulting large-scale unemployment?

In this context, those industries which have exhibited rapid technical change, have successfully increased productivity, and have at the same time reduced their labour force, are of particular interest. Thus the development of the post-war British gas industry warrants investigation: the quantity of gas sold has increased from around 2500 million therms a year in the early 1950s to over 15,000 million therms in the late 1970s; total gross output tripled between 1963 and 1974 alone; yet the labour force has been reduced from a peak of 147,967 in 1952 to 101,662 in 1976. That appreciable fall is made rather more spectacular if employment figures are broken down into manual workers and staff. When this is done, we find that during the same period the former category has been reduced from 112,121 to 40,674, while the latter has increased from 35,846 to 60,988. On the face of it these changes in the workforce appear to have taken place remarkably smoothly and it is of interest to enquire why this was so. Is it the case, for example, as some proponents of technical change claim, that an 'upgrading' of the workforce is an inevitable consequence of the adoption of more sophisticated technologies and that this can be accomplished through redeployment and re-training measures?[1] If not, then how were such large numbers of manual workers made redundant in an expanding industry with little apparent resistance? The

[1] This was a common argument in much of the writing on employment and automation which appeared in the late 1950s and early 1960s. See, for example [7].

TABLE 1. EMPLOYMENT IN THE GAS INDUSTRY (1950–1976)*

	1950	1954	1958	1962	1966	1970	1974	1976
Operatives	108,515	109,098	99,018	81,090	73,098	60,434	42,508	40,674
Administrative, technical and clerical	32,096	35,933	37,361	42,308	48,628	59,041	59,583	60,988
Total	140,611	145,031	136,379	123,398	121,726	119,475	102,091	101,662

* Annual Reports of the Gas Council and Gas Corporation.

answers to these questions may have some significance for the debates we are likely to be involved in in the near future.

THE BACKGROUND TO NATIONALISATION

During the inter-war years the gas industry expanded, but at a decreasing rate as the influence of competition from electricity slowly made itself felt. Inroads by electricity into the domestic market were to some extent compensated by increasing industrial sales, but the industry's performance in the immediate pre-war years was at best pedestrian. The structure of the industry militated strongly against greater competitiveness: gas was manufactured by the distillation of coal or was obtained from coke ovens, and was supplied by a host of independent undertakings, both company-owned and municipal,[2] while the research and development that was undertaken remained largely uncoordinated.[3] During the war, the industry suffered considerable physical damage to manufacturing, storage and distribution plant which seriously affected output and sales of gas. Moreover, the type of coal available, a crucial factor in gas manufacture, varied considerably and was frequently not of suitable quality, while the cost of both coal and labour increased steadily. Total output, nevertheless, grew throughout the war years,[4] particularly in those areas not subjected to heavy bombing, and this entailed the construction of new carbonising plant and the purchase of greater quantities of coke oven gas for distribution to consumers.

It was apparent, however, that more fundamental changes were essential to ensure continued survival in the post-war world, and in June 1944 a committee of enquiry was appointed to investigate the industry. The committee's report stressed that "the existing structure is restrictive of further progress" and recommended grouping the industry into large units, eliminating small plant, adopting mechanisation on a much greater scale, and encouraging research and development [8]. It was felt that these measures could best be achieved through nationalisation of the industry, and this was duly achieved with the passage of the 1948 Gas Act. The Gas Council, a directing agency, was formally constituted in

[2] The number of undertakings had reached a peak of 831 in 1913, and this had decreased to 703 in 1938.

[3] The Gas Light and Coke Company set up the London Research Station in 1924. Other undertakings which carried out research included the South Metropolitan Gas Company, and those in Birmingham, Liverpool and Bournemouth. The Institution of Gas Engineers provided a forum for the dissemination of much of this work.

[4] From 1,487 million therms in 1939 to 1,737 million in 1944.

TABLE 2. EXPANSION OF THE GAS INDUSTRY & PRODUCTIVITY (1950–1976)*

	1950	1954	1958	1962	1966	1970	1974	1976
Gas sold (million therms)	2,401.4	2,451.2	2,590.8	2,735.9	3,505.2	5,266.2	11,486.7	13,453.5
Number of consumers (thousands)	11,979.4	12,474.8	12,922.2	12,780.5	12,941.2	13,346.8	13,558.8	13,925.0
Number of employees/ million therms of gas sold	58.6	59.2	52.6	45.1	34.7	22.7	8.9	7.6

* Annual Reports of the Gas Council and Gas Corporation.

Vol. 7
1979

November 1948, twelve autonomous area boards were created, and on 1 May 1949 the gas industry passed into national ownership.

RATIONALISATION AND RESEARCH

By the time it was nationalised the industry was in a critical state. Much of the physical damage inflicted during the war was still apparent, plant was rapidly becoming obsolete, the price of coal to the gas industry was rising to the level of that charged to the domestic consumer, and there was a pronounced swing to the use of electricity for domestic purposes. More than anything, it was the price and quality of coal that gave cause for concern: it was essential to obtain supplies of high bituminous coal with low inert and high volatile content and with good coking qualities. (The generation of electricity, on the other hand, could be accomplished using vastly inferior coals.) Between April 1949 and March 1950, the gas industry carbonised over 25 million tons of coal and had to pay for it at two and a half times the price of pre-war years.

It was clear that any attempts to infuse the industry with new life would have to pay particular attention firstly to research and development programmes aimed at providing alternative gas manufacturing processes from coal, and secondly to changing the image of gas to enable it to compete effectively with electricity. In the former case the new Gas Council recognised from the beginning the need for action, and a research committee was set up to investigate possible courses of action. The Gas Research Board, which had been formed in 1939, provided a basis for expansion, although it employed only 38 staff in 1949 and had not been able to accomplish a great deal during the war years. Indeed, the overall lack of commitment to scientific and technological research within the industry was illustrated by the small number of employees engaged in such activities throughout the twelve area boards—only 401 with university degrees or professional qualifications in August 1949 [1]. As a result, two research stations were established, one based on the existing Fulham Laboratories of the North Thames Gas Board, the other in Birmingham under the auspices of the West Midlands Board, and at the latter a team continued work begun at Leeds University in the early 1940s on the problems of gasification of fuel oil and complete gasification of poor quality coal.

It was to be some time before this work bore fruit and in the meantime production continued along traditional lines. The grouping of the industry into area boards meant that a degree of rationalisation was possible and smaller gas works began to be closed and more modern plant introduced elsewhere. At first this had no adverse effects on the size of the labour force. Indeed, in the years immediately following nationalisation the number of manual workers and staff increased, and in certain areas (notably the Midlands where competition for labour was strong) deficiencies in numbers were a source of anxiety for a time. By 1952, however, the picture was beginning to change and the steady rundown of manual labour, which has characterised the industry since that date, was under way.

The nationalised industry inherited comprehensive national and regional negotiating machinery for manual workers and was pleased to report that it had found "a heritage of cordial and friendly relationships with the accredited representatives of all employees, which provided an excellent basis on which to build the new structure of negotiating machinery." [1, p. 34]. Under the terms of the 1948 Gas Act a National Joint Industrial Council was established, half the members of which were appointed by the appropriate Trades Unions[5] and half by the Gas Council, the intention being that the Joint Council should deal with the well-being of the industry, the improvement of working conditions, and "the attainment of improved output with a view to promoting the best interests of all those engaged in the industry" [1, p. 36].

The newly nationalised industry was fortunate in inheriting good labour relations and encountering a union attitude which was extremely co-operative in the face of a concerted programme of rationalisation. Indeed, it was almost impossible for union representatives to do anything other than accept the need for cutbacks in the labour force if the fragmented and

[5] The unions represented on the National Joint Industrial Council were the National Union of General & Municipal Workers, the Transport & General Workers Union, and the National Union of Enginemen, Firemen, Mechanics and Electrical Workers.

stagnant industry was to survive. Moreover, the existence of relatively full employment in the economy and the negotiation of an attractive redundancy scheme made the unions' decision not to oppose reductions in the labour force a less agonising one, and after reaching a peak of 112,121 in 1952, the number of manual workers steadily declined. The Gas Council was well pleased: "it is not too much to say that here, as in better service to customers, the progress which Boards have already made toward integration and the concentration of production in more economic units is being proved by the results." [2, p. 20]. By the mid 1950s between two and three thousand jobs a year were being lost and despite the Gas Council's official claim that "In the majority of cases it has been possible to offer appropriate alternative employment", [3, p. 26], it was generally the case that redundancy was the path most often chosen by those whose jobs were disappearing.

At this stage of the industry's development, technical change played almost no part in the loss of jobs, and the Gas Council replied to the Board of Trade's enquiries concerning its use of automation in 1957 that the greatest opportunity for increasing productivity lay less in technical change and more in concentrating production "in the larger works and closing down the small works which are mostly hand-operated." [4, p. 12]. At the same time, however, developments were taking place on the research front which were to have an even greater impact upon the industry's labour force. As the price of coal continued to rise, so the need for alternative methods of gas production became greater: by the end of 1955, increases in coal prices were costing the industry annually almost £50 million more than would have been the case at pre-nationalisation prices. Consequently, the industry's research programme was stepped up and given the three aims of improving the efficiency of existing methods of manufacture; developing new processes of manufacture and new sources for the supply of gas; and developing greater efficiency in the utilisation of gas and other products of gas making. In particular, oil hydrogenation was, by the mid-1950s, regarded as a promising process: refinery gases and light distillates were in plentiful supply because of the development of oil refining, and the appropriate techniques were convenient and provided a quick response to a sudden increase in gas demand. Early experiments had proved successful and trial production on a commercial scale had been carried out since February 1953 at Sydenham, using the industry's own process, and at Stafford, using the Onia–Geigi process, since May 1954. After the programme of rationalisation, the stage was set for a second major upheaval within the industry.[6]

OIL GASIFICATION: A MAJOR INNOVATION

As a greater commitment was made to obtaining gas from oil, so the composition of the industry's labour force began to change. Recruitment was from significantly different quarters: there was "a need for draughtsmen, chemists and technicians" and it was hoped that the industry could attract its share "of the technical and scientific candidates likely to come in increasing numbers from the universities and technical colleges." [4, p. 25]. As the new class of technicians grew, the number of manual workers decreased at an even faster rate. Some found employment in the new oil based plants, where physical conditions were better, but by 1962 over 30,000 jobs had been lost in the previous 10 years, due, in the Gas Council's words, to "continued integration, adoption of modern automatic methods of production and distribution, and to centralised administration." [5, p. 25].

While overall employment was declining, the industry began to show heartening signs of recovery and growth. Indeed, for an enterprise that many felt to be moribund in the late 1940s and early 1950s, the change was remarkable. The annihilation of gas by electricity, that was confidently predicted after the war, had not occurred, and the 'dirty and dangerous' image of gas that had plagued the industry was replaced by one more in keeping with the 'white heat of the technological revolution' that lay around the corner. Much of this successful recovery was attributable to a massive investment in management recruiting and training and to important work carried out on the domestic utilisation of gas appliances which

[6] The Onia–Geigi process was followed in 1962 by the adoption of the ICI catalytic high pressure reformer which gave the industry a very cheap, sulphur-free town gas.

Vol. 7
1979

TABLE 3. CHANGES IN THE SOURCE OF GAS. (1961/62–1967/68)*

	1961/62	1962/63	1963/64	1964/65	1965/66	1966/67	1967/68
Gas made (million therms)							
1. Coal gas	1704.9	1684.0	1637.7	1461.8	1269.1	1152.7	976.8
2. Lurgi gas (lean)	24.2	32.1	51.9	75.3	69.0	66.8	64.5
3. Water gas	436.4	534.2	454.4	419.7	379.1	255.8	159.7
4. Oil gas	76.8	132.6	197.6	406.0	692.3	1151.7	1660.2
5. Producer gas & other	20.8	21.0	15.6	10.6	6.8	5.6	3.6
Gas bought (million therms)							
6. From NCB coke ovens	274.7	254.9	232.9	222.4	219.8	213.0	202.2
7. From other coke ovens	237.2	220.0	227.8	225.0	216.5	196.2	192.3
8. Oil refinery gas	154.1	222.1	258.5	279.6	302.2	309.8	328.5
9. Liquified petroleum gas	40.1	84.0	155.0	286.8	371.8	448.0	409.6
10. Natural gas & other	13.9	20.8	31.5	133.4	313.7	314.8	626.7
11. Direct natural gas							13.4
Total gas available (million therms)	2983.1	3205.7	3262.9	3520.6	3840.3	4114.4	4637.5

* Gas Council Annual Report for 1968, p. 132.

resulted, for example, in the efficiency of gas fires being increased from around 30% to over 60% in little more than a year. Underlying these developments, and to a large extent consolidating them, were the benefits accruing from the new oil-based gas manufacturing techniques, in particular a higher quality gas and, perhaps most important of all, stable prices as reliance on coal diminished.[7] Table 3 illustrates the changing source of gas in the period.

During this period of growth the vast majority of workers who were displaced continued, as previously, to choose redundancy rather than redeployment or re-training. Yet this was no longer such an obvious choice as it had been in the past. The industry was, after all, now expanding rather than fighting for survival and the possibilities for remaining within it appeared considerably greater. The explanation for the willingness to accept redundancy lies chiefly in the past success of negotiated compensation schemes[8] and in a widespread tendency among manual workers and union officials to see the industry as a 'redundancy industry' after the events of the 1950s. As a result, union negotiators pressed harder for better redundancy payments, longer periods of notice and the like, but offered no challenge to the basic management strategy of reducing numbers. A considerable momentum was built up around the notion of redundancy, and the majority of manual workers expected to be offered those terms (and even better ones) which were presented to their predecessors. Again, a situation of relatively full employment in the country at large contributed to this willingness to leave the industry.

In retrospect it would have been possible for union officials to advocate more strongly retraining and redeployment measures, yet the strong preference for redundancy among their members (expressed both at regional levels and at national conferences) would have required a major effort of persuasion. Besides, union officials themselves had become extremely skilled at negotiating attractive redundancy agreements and were thus constrained by the expertise they had developed.

Since it was necessary to keep plants operating until the last moment, workers were informed of impending closure and offered attractive redundancy payments conditional upon their staying until discharged by management. It was made clear to employees that leaving the industry prior to the closing down of the operation would prejudice their entitlement to redundancy pay. This last minute closure of works presented little opportunity for retraining or absorbing large numbers of workers. Moreover, the short time allowed employees to choose between possible redeployment and redundancy coupled with uncertainty surrounding the nature of alternative

[7] Between 1961/62 and 1967/68 the quantity of coal used in gas manufacture decreased from 22,213,000 tons to 13,512,000 tons; that of oil increased from 866,000 tons to 4,605,000 tons.

[8] The first compensation scheme came into effect in May 1959 and was superceded by a second in January 1964.

employment meant that cash payments appeared particularly attractive. Even in those areas where trial periods of alternative employment were allowed, the tendency was for workers to choose redundancy, producing, as one union official admitted, "a result highly satisfactory" to management, unions, and, at the time, to workers.

THE TRANSITION TO NATURAL GAS

A striking indication of the changes that had been taking place in the industry was provided in 1964 when consumer service became the largest single category of employment. At the same time developments were afoot which were to reduce the number of those involved in gas manufacture even more drastically than had hitherto been the case: the search for North Sea gas.

Exploratory drilling for natural gas had been undertaken in various parts of mainland Britain as early as 1952, although the industry's first real large-scale experience of natural gas was provided later in the decade when proposals were made for importing liquified methane from Algeria. In November 1961 an agreement was signed for the shipping of methane equivalent to 350 million therms a year, for 15 years, to terminals on Canvey Island. As part of this project a backbone main was constructed from London to Leeds, foreshadowing developments which were soon to become an integral part of the industry's activities.

After the Dutch discovery of North Sea gas in 1963 it became clear that the industry could be revolutionised once more. In retrospect (although it was not recognised at the time) it is clear that a transitional stage between traditional coal gas manufacture and natural gas was an essential prerequisite for successful exploitation of the new resources. Fortunately, the modernisation of the industry that had accompanied the large scale gasification of oil provided just such a bridge. As a result the industry had become more scientifically minded, and had acquired a considerable body of technical skill which was to prove invaluable in making use of the more sophisticated technologies demanded by natural gas. The Gas Council was aware of the challenge in terms of manpower: "With its present expansion and the need to plan for the introduction of natural gas from the North Sea, the industry offers excellent opportunities for capable, intelligent, enthusiastic and responsible people. Recruitment efforts in many fields are being stepped up to meet the keen competition from other employers." [6, p. 41].

For the industry's labour force, and the trades unions, the implications of these fundamental changes were unclear. It was apparent that jobs in gas manufacture would decline at an even faster rate, but on the other hand there would be substantial opportunities for employment within the distribution sector and in the conversion of an estimated 30 million appliances. At this point, with a rapid expansion of the industry in prospect, a major shift in union policy could have taken place. Significantly it did not occur, and the policy of seeking redundancies remained unchallenged, the labour force continuing to express a preference for better redundancy payments.

In fact, much of the new distribution and conversion work was undertaken by contractors and very few employees of the industry transferred to the new programmes. The extent to which it would have been possible to persuade more workers to remain within the industry is debatable. On the one hand they had pension rights and the like to protect; on the other the work involved in distribution and conversion was very different from their previous experience and the contrast between factory-like shift work and the new employment, demanding flexibility in both hours and mobility, may have been a major obstacle.

Since the advent of natural gas, the impressive progress of the industry has been maintained. In the first 5 years of this decade, total supplies of gas available doubled to over 14,000 million therms; industrial gas sales increased by three and a half times; and employment of manual workers declined by a further 14,000. Even the total of administrative, technical and clerical staff has stabilised, thus bringing an end to the growth which had characterised this section of the industry since nationalisation. Indeed, a comparison of the fortunes of manual workers and staff reveals some interesting features, not least that while the former have declined by almost two thirds since 1952, the latter have almost doubled in number. In part this reflects a reclassification of manual jobs as staff posts—meter readers

Vol. 7
1979

and collectors for example—while much of the increase can be attributed to the expansion of administrative staff numbers as the industry has grown, and to the greater commitment to scientific and technical research associated with an industry which has experienced rapid technological change. Nonetheless, it does appear that staff have grasped more readily than manual workers those opportunities for redeployment and retraining that have presented themselves. Many gas engineers, for example, previously employed in gas manufacture were well aware that their skills would find little application outside the industry and were therefore anxious to find alternative employment within it. This contrasts markedly with manual workers who felt certain they could obtain substantial redundancy payments and secure a comparable job elsewhere. Clerks, shorthand typists, and machine operators were all easily transferred within the industry since their skills were in short supply. Only at the supervisor level were problems frequently encountered, and significantly many of those involved had come from the shop floor and were in many respects the staff closest to the manual workers.

CONCLUSION

The question remains as to how many manual workers could have been kept within the industry if re-training policies had been pursued more vigorously. Since the gas industry is now having to recruit large numbers of people for distribution work—up to 5000 workers are needed in the near future—it would seem that the potential existed for redeployment on a fairly large scale. Other posts could also have been filled by displaced production workers, notably in maintenance, store-keeping and similar jobs, and in administration. Given the strong distinction between manual and non-manual work, the last of these was perhaps unlikely to account for many displaced production workers, although union representatives insist that the other categories were capable of absorbing a considerable number. The difficulties of actually persuading men to re-train or to undertake work of a different nature would remain. Certainly, attempts to re-train workers as fitters (a high status job within the industry) met with very limited success, requiring as it did a long training in parallel with the apprenticeship scheme and examinations at the end which deterred older men from taking part. It comes as no surprise that generous redundancy payments should have proved such an attractive alternative.

The mismatch in terms of skills and worker expectations between old jobs and potential new ones emerges as an important issue in a consideration of the gas industry's successful shedding of labour. Thus while the changing composition of the industry's entire workforce appears to support the 'upgrading' thesis mentioned in the introduction to this paper, such changes owed little, if anything, to retraining and redeployment within a context of technological change. Rather, it has been the case that recruitment to new staff and technical posts has taken place from outside the industry, while displaced production workers have left the industry, often in the expectation of finding similar work elsewhere. Some of the problems of re-training and redeployment were of course peculiar to the gas industry—for example, an ageing workforce which was more favourably disposed towards accepting redundancy payments than towards a different type of employment. Yet some of the difficulties involved in re-training appear to have a wider significance. In particular, the problem of presenting to those displaced work of a quite different nature remains of central importance. Industries adopting new technologies and wishing to shed labour in future will not be able to do so in an economic climate as favourable as that enjoyed by the post-war gas industry nor can they expect such willing co-operation from trades unions. With a dearth of alternative employment it will be yet more important for both management and unions to cultivate an atmosphere within which proposals for re-training can flourish. In a rapidly changing technological world it will have to be made an integral part of work experience rather than a special event.

ACKNOWLEDGEMENTS

The author would like to acknowledge the contribution to this paper of Professor Ernest Braun, Technology Policy Unit, University of Aston. Mr John Doran, ex-chairman of the East Midlands Gas Region, and Mr John Edmonds, National Industrial Officer of the General and Municipal Workers Union, kindly provided much valuable information.

REFERENCES

1. Gas Council (1951) *First Report & Statement of Accounts for the Period 30 July 1948 to 31 March 1950.* HMSO, London.
2. Gas Council (1953) *Annual Report & Accounts for the Year Ended 31 March 1953.* HMSO, London.
3. Gas Council (1955) *Annual Report for the Year Ending 31 March 1955.* HMSO, London.
4. Gas Council (1957) *Annual Report for the Year Ending 31 March 1957.* HMSO, London.
5. Gas Council (1962) *Annual Report for the Year Ending 31 March 1962.* HMSO, London.
6. Gas Council (1966) *Annual Report for the Year Ending 31 March 1966.* HMSO London.
7. Goodman LL (1957) *Man & Automation.* Penguin, Harmondsworth.
8. Ministry of Fuel & Power (1945) *The Gas Industry. Report of the Committee of Enquiry.* HMSO, London.

ADDRESS FOR CORRESPONDENCE: *Dr Russel Moseley, Technology Policy Unit, University of Aston in Birmingham, Gosta Green, George Alexander Building, Birmingham B4 7ET.*

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 2. pp. 113 to 117
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0401-0113$02.00/0

Budgets for Integrating Decentralized Investment and Production Planning

RICHARD M BURTON

Duke University, Durham, North Carolina

WILLIAM W DAMON

Vanderbilt University, Nashville, Tennessee

(*Received October 1977; in revised form November 1978*)

Planning for an economic enterprise can be dichotomized into short-run production planning and longer-run investment planning. Usually these problems are treated as if they were separate, if not independent. This paper briefly reviews the separate approaches to optimal production decision making and investment planning, 'fusing' these models in order to consider the two issues simultaneously. The resulting 'fused' model is used to illustrate several difficulties which result from an intuitive synthesis of the independent solutions of the production problem and the investment problem. An integrated model is presented representing a centralized simultaneous solution for decision variables from the two functional fields. The paper compares and contrasts the synthesis of separate functional models to the decomposition of a simultaneous model of those functional areas. A result of this comparison is a theoretical justification for operating budgets and revenue targets as organizational mechanisms for achieving coordinated plans among decentralized planning units. Further, the set of conditions are identified under which the two approaches to simultaneous decision making are equivalent.

PRODUCTION AND INVESTMENT PLANNING

INITIALLY, the production planning problem is presented as a linear program and the net present value investment decision rule is reviewed. These separate models are then synthesized into a 'fused' model of production and investment planning.

Consider the following single period linear programming production planning model:

$$\begin{array}{ll} \max & \Sigma\, c_j\, x_j \\ \text{subject to} & \Sigma\, a_{ij}\, x_j \le b_i \\ & x_j \ge 0 \end{array}$$

where:

Σ stands throughout for summation over all values of j, unless stated otherwise

x_j = the amount of output of the j^{th} product to be produced (a decision variable).

b_i = the quantity of the i^{th} input which is a fixed factor for the short run planning problem.

a_{ij} = the quantity of the i^{th} fixed factor required per unit of the j^{th} product.

c_j = the contribution to the cash flow (selling price minus variable costs) of one unit of the j^{th} output.

Let π_i represent the optimal value of the dual variable associated with the fixed input b_i. Then π_i measures the increment to cash flow if one more unit of the first capital good is made available to production while all other fixed resources are held constant, provided the primal solution remains feasible.

It is assumed that an additional unit of fixed input b_j is available not only in the current period, but also in future periods. However, the input's availability is subject to deterioration

over time at a rate of $(1 - \alpha)$ per period. Thus, subsequent production planning problems can be represented by the same linear programming formulation by adding a time subscript to all parameters and decision variables (so that b_{it} represents the anticipated level of capital good i available in the subsequent period t with the shadow price π_{it}).

The acquisition of an additional unit of the first fixed-input factor in period 1 (and implicitely α^t units in period t) can be evaluated using capital budgeting techniques. If each unit of factor 1 costs g_1, the net present value of an investment in one unit is:

$$NPV = \sum_{t=1}^{T} \frac{\alpha^{t-1}\pi_{it}}{(1+r)^t} - g_1$$

The decision rule is to add to resource 1 if the NPV is positive. This decision rule will be referred to as the fused model since it fuses the optimal marginal cash flows, or shadow prices, from production to the Net Present Value investment decision rule.

Several shortcomings of this fused model as a practical decision rule are quite apparent. Letting e_i represent the level of investment in resource i, the decision rule implies that optimally either e_i is zero ($NPV < 0$) or e_i is infinite ($NPV > 0$). Second, it is based on the assumption that for any level of change of b_{it} that the shadow prices are constant. Clearly these first two issues are closely related. A third difficulty is that once an investment in capital good 1 is determined, the revised production planning problem will provide new dual variable values for capital good i if the optimal basis changes. Thus, different sequences of capital good decisions may determine different investment decisions. Further, the fused models fails to identify an organizational process to achieve a solution.

The first two shortcomings could be dealt with by recognizing that π_{1t} is a function of e_1 and by replacing π_{1t} in the NPV decision rule with the monotone decreasing step function (analogous to the marginal revenue curves derived in Vander Meulen [14]) which relates π_{1t} to e_1. An optimal value of e_1 can then be determined by increasing e_1 as long as the marginal increment to present value remains positive. However, this analysis would not resolve the issue of the appropriate sequencing of investment decisions.

A MULTI-PERIOD PRODUCTION AND INVESTMENT PLANNING MODEL

In this section, we present a mathematical programming model designed to provide the simultaneous determination of the short-run production variables and the long-run investment variables. The model is a linear programming model structured such that a budget decomposition (see [3]) is a suitable means for its solution. Linear technology and perfectly competitive product and factor markets are assumed.

We have chosen to ignore the possibilities of external debt or equity funding of investment decisions in order to focus on the interactions between production planning and investment decision making endogeneously. Morris [12] discussed these external financing issues assuming exogenous cash flow estimates of investments. Our objective is to determine the cash flow implications of investments in productive resources within the simultaneous model based upon the optimal utilization of such investments in the production of the firm's outputs.

We can write the model as:

$$\text{Max} \quad \sum_{t=1}^{T} D_t/(1+r)^t$$

Subject to (i)

$$\sum a_{ijt} X_{jt} \leqslant b_{it} = \alpha b_{i,t-1} + e_{it}$$

(ii) Terminal Cash = Initial Cash + Revenues − Production expenses − Investment expenditures − Dividends

The objective function is the present value of the cash dividends, D_t, to the firm's owners. The first set of constraints, (i), deals with physical limitations of production inputs in each time period (analogous to the production model given previously). Also within (i) we represent the level of capital equipment in time period t as the amount that was available in the previous time period adjusted by a survival factor, α, plus the amount purchased at the beginning of the time period t. The constraints in (ii) deal with the cash position of the firm and essentially equate the sources and uses of cash in each time period.

The set of conditions indicating the technological depreciation of available facilities from period t to period $t + 1$, and the potential for

Vol. 7
1979

cash carryover, are the only restrictions which tie the planning problem together, such that it is not trivially decomposable into a planning problem for each period individually.

Writing each b_{it} in terms of initial stock and subsequent investments, it is obvious that production decisions in any given period depend upon the investment decisions of all previous periods.

$$b_{it} = \sum_{k=1}^{t} \alpha^{t-k} e_{ik} + \alpha^t b_{io}$$

Looking forward, we also see that today's investment decisions create constraints on production decisions for future periods.

DECOMPOSITION OF THE MULTI-PERIOD MODEL

There are two basic approaches to decomposition for this type of model. One is the price coordinated approach developed by Dantzig and Wolfe (D–W) [17, 8] and later interpreted for the firm by Baumol and Fabian [2]. The other method is a budget[1] approach as developed by Kornai and Liptak (K–L) [10] for national planning, and later interpreted for the firm by Burton, Damon and Loughridge [3]. A modified budget method which assures finite convergence was developed by Ten Kate [13].

The choice of both the form of decomposition and of the partitioning of the decision variables into 'divisions' of the firm are essentially the issues of organization design. Each different partition of the decision variables represents an organizational form which has an implicit information system which is derived from the mechanics of the information flows associated with each iteration of the planning process. We choose to ignore many of the 'natural' partitions (product line, geography, common resource utilization, etc.) in favor of a temporal decomposition (or decentralization). The rationale for this choice is that a temporal decomposition is a generalization of the fused model presented earlier. A further rationale is that it provides insight into the decentralization of the firm into the functional areas of production planning and investment decision making.

In order to relate the fused model and the simultaneous model we have selected a partitioning of the decision variables which would require independent determination of each period's production decisions, the vector $\mathbf{x}_t$, and an overall coordinator responsible for non-production decision variables, including investments. The separate period production plans must be coordinated due to the coupling requirements that exist among the remaining variables. The budget and investment organizational unit is responsible for setting capital input levels (b_{it}), investments (e_{it}), dividends (D_t), and is clearly the coordinating unit for the production planning periods as shown in Fig. 1. Figure 1 demonstrates the type of calculation each organizational unit makes and the necessary information flows to coordinate the decisions of these units.

The budget form of decomposition for this model requires the definition of an operating budget, B_t, and a revenue target, R_t, for each period. The created budget constraints, B_t, and revenue requirements, R_t, for each period guide the choice of the production plan. They drop out if the period production models are substituted back into the central coordinator's model to reconstruct the original simultaneous model. Thus, they do not appear in the original simultaneous model. This decomposition approach to the solution of the simultaneous model provides a theoretical justification for introducing operating budgets and revenue targets as coordination mechanisms for the decentralized solution of the firm's planning problems.

For this decomposition, the budget coordination model and each production planning model are formulated as linear programs. The production models are the original production model with operating budgets, and revenue requirements added. The budget coordination model is similar to the simultaneous model, except the production variables are replaced by their aggregate effects summarized in the operating budgets and revenue targets.[2]

[1] Budget is used in its general meaning to denote monetary allocations and requirements as well as non monetary allocations, such as the number of machines available.

[2] The Kornai–Liptak decomposition algorithm will maintain primal feasibility at every iteration if the associated game is regular. The subproblems' feasibility could be guaranteed by allowing the subproblems access to artificial external financial markets at artificially high rates. We refrain from introducing these markets since it adds considerable notational complexity without changing the central argument.

Further, the K–L approach uses a weighting procedure of past provisional budgets and current solution in the information flows. For a linear model, such a scheme is required to obtain convergence.

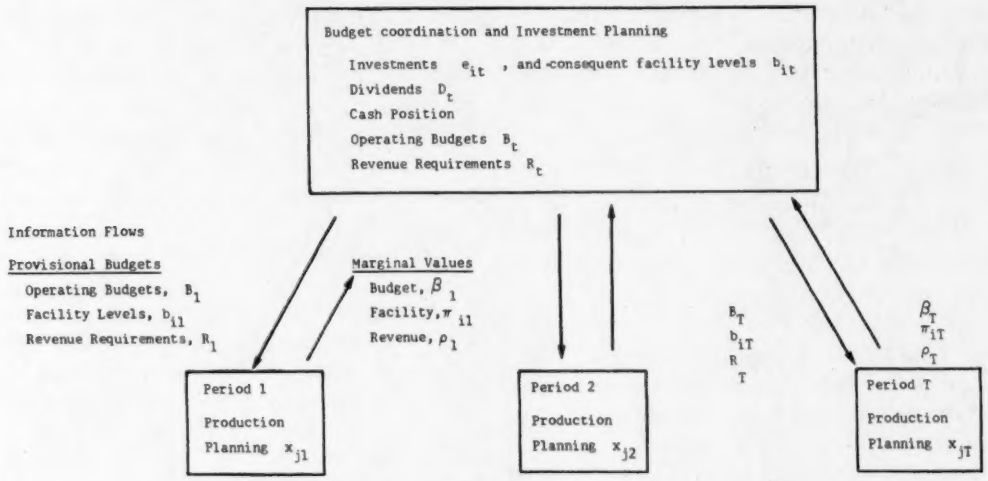


FIG. 1. *The planning scheme.*

An important feature of this solution procedure is that at the termination of the algorithm each organizational unit can, and must, implement the solution it has calculated on the final iteration. This feature is important because it contrasts with the transfer price decomposition algorithm in which the solution to be implemented is a weighted average of previous solutions rather than the last solution recommended by the subsystems. Behaviorly, this resource allocation approach is more viable. (See [4] for a more complete discussion).

COMPARISON OF THE MODELS

The 'fused model' is the net present value decision rule applied to the investment in one more unit of one of the productive resources (arbitrarily resource 1). The 'fused model' is:

$$NPV = \sum_{t=1}^{T} \frac{\alpha^{t-1}\pi_{1t}}{(1+r)^t} - g_1 \quad \begin{matrix} > 0 \text{ accept} \\ \leqslant 0 \text{ reject} \end{matrix}$$

The application of this model presumes that the most recent schedule of investments will be undertaken so that subsequent production plans may reflect the resultant expansion of the scarce resources. With each period's optimal use of the resources established, the shadow prices in each period, π_{it}, are known.

We apply analogous assumptions to the decomposition of the simultaneous model and consider the resultant problem of the budget coordinator. To mirror the assumptions of the fused model each production plan must have a sufficient operating budget, B_t, and a non restrictive revenue target, R_t. This can be achieved by setting both such that they are nonbinding, and thus their shadow prices (β_t and ρ_t) are equal to zero. With the current investment plan which is updated as a starting point and treating all planned investments as fixed except for the first period (analogous to the fused model), and letting C_{1t} represent the T–t period rental value per unit of productive resource 1, the central coordinator's optimization problem can be written as:

$$\max e_{11}\left[\sum_{t=1}^{T}\left(\left(\frac{1}{1+r}\right)^t \alpha^{t-1}\rho_{1t}\right) - C_{11}\right]$$

This expression for the first iteration of the central coordinators problem is identical to the fused model decision rule presented earlier and thus the optimal value of the function is obtained when the investment quantity, e_{11}, is either zero or infinite.

This equivalence of the two models is a measure of the crudeness of the fused model, not a justification of the simultaneous model. In fact ideally the first iteration of the decomposition algorithm should not be the iteration used above for purposes of comparison. The viability of this decomposition approach would rely upon the use of the previous period's tentative investment plans as the first iteration of the algorithm for the current period. This initial solution would be improved upon by the use of the operating budgets and revenue targets as coordinating mechanisms. Since in practice only a few iterations would actually occur before a plan would have to be imple-

mented, starting with last period's tentative allocations and budgets 'rolled' forward one period would assure a good starting point for the revision, or iteration, process.[3]

CONCLUSION

This paper presents a fused model of the net present value decision rule for investment decisions in which the marginal impact upon future cash flows is measured by shadow prices from a linear programming formulation of the production processes of the firm. The shortcomings of the fused model are discussed. A linear program is presented which represents the simultaneous problems of production planning and investment decision making. Using a budget decomposition, it is demonstrated that one possible first iteration of the decomposition algorithm leads to an investment decision problem identical to the fused model. Operating budgets and revenue targets are introduced to coordinate consecutive production planning problems, and to overcome the shortcomings of the fused model. These budgets are the organizational mechanism by which the firm achieves a coordinated plan for production and investment where functional planning units are organizationally decentralized.

[3] Computational experience on related models suggests that budget decompositions using historical data to start the planning scheme are effective and yield good solutions in a few iterations. See Burton and Obel [5].

REFERENCES

1. Baligh HH & Burton RM (1976) Organization structure and cooperative market relations. *Omega* **4**(5) 583–593.
2. Baumol WJ & Tibor F (1964) Decomposition, pricing for decentralization and external economics. *Mgmt Sci.* **11**(1) 1–32.
3. Burton RM, Damon WW & Loughridge DW (1974) The economics of decomposition: resource allocation vs. transfer pricing. *Decis. Sci.*
4. Burton RM & Obel B (1977) The multilevel approach to organizational issues of the firm. *Omega* **5**(4) 395–414.
5. Burton RM & Obel B (1978) Empirical results of the relative efficiency of the price, budget and mixed approaches for decentralised planning. Technical Report, Odense University, Denmark.
6. Damon WW & Schramm R (1972) A simultaneous decision model for production, marketing and finance. *Mgmt Sci.* **19**(2) 161–172.
7. Dantzig G & Wolfe P (1961) The decomposition algorithm for linear programming. *Econometrica* **29**(4)
8. Dantzig G & Wolfe P (1960) Decomposition principles for linear programs. *Ops Rs.* **8**(1) 101–111.
9. Hannsmann F & Hess S (1960) A linear-programming approach to production and employment scheduling. *Mgmt Technol.* (1).
10. Kornai J & Liptak T (1965) Two-level planning. *Econometrica* **33**(1) 141–169
11. Loughridge DW, Damon WW & Peterson DW (1976) The coordination of short-run decision-making with long-range planning. *Omega* **4**(6) 673–684.
12. Morris JR (1975) An application of the decomposition principle to financial decision models. *Jl Fin & Quantit. An.*
13. Ten Kate A (1972) Decomposition of linear programs by direct distribution. *Econometrica* **40**(5) 883–898.
14. Vandermeulen DC (1971) Linear Economic Theory. Prenticehall, Englewood Cliffs, New Jersey.
15. Weingartner M (1967) Mathematical Programming and the Analysis of Capital Budgeting Problems. Markham Publishing, Chicago, USA.

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 2. pp. 119 to 127
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0401-0119$02.00/0

On the Nature of Problems in Consulting Practice[1]

COLIN EDEN

DAVID SIMS

University of Bath, UK

(*Received August 1978; in revised form October 1978*)

OR consultants who wish to help clients with complex policy issues need to find ways of taking account of personal and illegitimate features of a client's understanding of his situation, as well as the technical features which are usually considered. This paper is an attempt to discuss the *practice* of operational research in ways which reflect its political complexity, rather than the *techniques*, which are more commonly described in the published papers and texts. The discussion leads to the suggestion that practitioners could usefully devote some energy to reflecting upon the processes used in effective problem solving.

INTRODUCTION

FOR MOST of its life OR has been characterised by a perspective based upon the notion of problem *solving* from a so-called inter-disciplinary understanding of the social world. The skills and expertise of the OR man thus enable him to offer solutions to a wide range of different problems such as manpower planning, production scheduling, and marketing tactics.

There is a widespread belief that whilst problems may occur in many guises requiring different ways of gathering data and translating it to a manipulable form, the general approach to solution remains the same in all cases. Implicit in this is another belief, that the solution of a problem may usefully be regarded quite independently of the formulation, data gathering and solution implementation for the problem.

Underlying both these beliefs is another belief that problems are entities in their own right. This seems to stem from a 'systems' view of the world [12]—that is they exist outside of a person and can thus be discovered and described as a result of the OR inquiry process. If only the operational researcher can separate off the 'real' problem from any other (and by implication 'unreal') problems which may be presented, then so long as he is given adequate access to the facts he can proceed to modelling and solution. In this way it is possible for the methods of OR practice to remain intact by implicitly ignoring the complicated qualitative experience and wisdom of those who have endured the problem environment for much of their occupational lives but who unfortunately do not have the skills to construct 'respectable' science–rational models of their complex environment.

Traditional methods of science–rational analysis show an understanding of the social world which people have often commended to each other, but which we would suggest no one finds to be a useful way of operating for himself. It may be that we all believe that we are the only people with complex problems, and

[1] This paper was presented at 'The King is Dead: Long Live the King'; One-day conference organized by the National Events Committee of the Operational Research Society. December 1977.

that others just suffer from muddled thinking; for this or some other reason there is a misunderstanding of the relationship between problems and the persons who have them.

Several studies have demonstrated the disappointing consequences of this framework for problem solving—Brewer [11] conducted an extensive analysis of the effect of OR in the solution of urban problems and discovered an over-riding disillusionment where apparently none of the parties associated with the activity (public, politicians, civil servants, and analysts) appeared to recognise a payoff to anyone either overt or covert. In this paper we shall argue that some disillusionment is an inevitable consequence of the underlying paradigm guiding OR practice. This paradigm is the result of an over emphasis on a particular view of science which has led implicitly to the dominant guiding principles being derived from the Systems Frame of Reference [10]. We shall also seek to demonstrate that the explicit recognition of alternative frames of reference could potentially decrease the likelihood of disillusionment on the part of client or consultant.

Our argument does not derive its impetus from a belief that current OR practice is always poor or unsuccessful but more from the presumption that a more sound theoretical basis for understanding and reflecting upon the process of effective consulting will inevitably lead to a better profession. The perspective we shall present is driven by an analysis of successful consultant practice and an attempt to map the results of this analysis into an appropriate theoretical framework.

THE REALITY OF PROBLEMS AND THE PROBLEMS OF REALITY

There is a major difference in emphasis between discussion within the profession and the behaviour of the OR consultant working within a social organization. In the first instance we see, within textbooks, journals, project reports, proceedings of conferences, an attention to the problem—its characteristics, structure, content—that is to say our attention is directed to an objective reality [17], a system of interacting variables that as a consequence of manipulation could be made to behave differently. The description of the problem is implicitly contained by the form of the model which is used for its solution, a solution is discovered (usually through mathematical/numerical manipulation) which will enable the system to operate in a preferred manner. In this way we are led to believe in the reality of an objectively defined problem which is assumed to be self evident to all intelligent people; those who do not see the facts of the matter are therefore regarded as recalcitrant. However when we study what is going on in the process of behaving as a consultant we see a part of OR practice which we apparently feel unable to discuss, reflect upon, or theorize about. We see a complicated drama unfold which involves power, influence, negotiation, game playing, organization politics, complex social relationships with real people not merely office holders [26]. In this environment problems are not self evident at all; under the guise of the same problem title each actor sees a reality which is unique to him, that which comes to be known as the *real* problem for the consultant depends then upon his *own* reality and that belonging to those actors to whom he chooses to listen.

It seems almost trivial to note that it is not possible to talk meaningfully about a problem without ascribing an owner or owners to it, because different persons will experience different problems in what might be thought of as the same situation. We do not mean to suggest that problems are wholly and irreducibly individual, since it is clearly possible for a group to own a problem as a consequence of negotiation [28]. However, for reasons that we shall explain later, we believe that too little attention has been paid to this process of negotiation. For example, the event 'shop steward calls strike meeting' may be an event about which we all know, however for the purpose of understanding the nature of problems we need to discover how this event is *interpreted.* A consultant may *define* a problem based upon an interpretation of the event as the predicted consequence of a poor incentive scheme, the managing director sees it as a storm in a tea cup and of no real consequence, to the industrial relations director it is the possibility of his being replaced, and to the local district union official it is the shop steward asserting his authority so that he may take over as district organiser. Each of these individuals could become a client or indeed all of them may wish to define themselves as a client group.

Vol. 7
1979

Such discrepancies in the way that problems are described may be attributed to wilful misrepresentation, but we believe that some discrepancy of this kind is the natural result of the cognitive process. The person sees his world with the aid of a unique mental framework; the world around him is much too complicated for him to perceive every part of it and every possible combination of parts, and a framework is necessary for him to make sense of events and objects in his environment. It follows that problems can only be the outcome of *making sense* of the reality the person has constructed through the use of his mental frame of reference; if after the use of his frame of reference he has some unease or dissatisfaction which he does not know how to deal with, we say he has a problem [23]. If after the use of his framework he does not feel any unease or dissatisfaction then no matter how appalling his situation or how golden his opportunity may appear to an observer he does not have a problem.[2]

How do we account for this process of problem construction theoretically? Following theories from the sociology of defining situations[3] and the psychology of cognition,[4] problem *definition* can be explained by the way in which each of us has developed a unique schema (mental framework) for interpreting the events around us. Perception and interpretation are ordered by the contents of this framework which include systems of values and beliefs.

For most of the problems which operational researchers are called upon to solve, however, this is too simple a view. When a person in an organization has defined a situation as being problematic, he may be able to go ahead and formulate and solve the problem for himself, but in this case no operational researcher is involved. He may be able to call in a consultant for help without reference to his peers and superordinates, but even if he has the right to do this he may not be willing to accept the consequences for himself within the organization of doing so, which may be seen as high-handed, as suspicious manoeuvring or simply as an admission of incompetence.

So we may expect that, most commonly, problems presented to an operational researcher will be the result not only of one man's schema, but also of a process of negotiation and redefinition within a group. Let us consider an initial problem definer bringing his problem to his work team. However he has formulated the problem for himself, the way he describes it to the team will be influenced by the way he wishes to present himself to that team. If he wants them to think of him as competent he is unlikely to present a problem in such a way that it could imply he has acted incompetently, even if this were his private belief about the genesis of the problem. If he wants the team to take on the problem as being its own, rather than just sanctioning him to get help with it, he will present the problem in such a way as to make it apparent that there are gains to be had or losses to be averted for other members of the team by solving this problem. He may seek to show some member of the team that a solution to his (the initial problem definer's) problem would also solve some different problem which he believes this member to be experiencing. There often seem to be fashions for a type of problem in a particular team or organization, and he may define his problem to be in line with other problems which seem to be being experienced in the team at that time. Or, maybe he needs to express concern and commitment about some problem being stated by another member in the hope of getting some concern and commitment about his problem in return.

It becomes apparent that the process of problem definition that is an inherent part of the consultancy role does not essentially involve modelling an objective reality but does entail identifying whose reality is to be attended to. It is a matter of *choice* for the consultant. This choice does not usually result in the exclusive attention to one individual rather than others, but rather a choice about some mixture which will often reflect the consultant's understanding of power and influence in the group. We maintain that this choice is very often not made with the care that is required if the consultant wishes to satisfy the needs of any particular client group or simply

[2] This type of perspective is accounted for by sociologists under the heading of phenomenology: see for example Bittner [7].

[3] See McHugh [24] and Ball [5] for a discussion of the implications of 'defining situations' on understanding social behaviour.

[4] For an excellent introduction to cognition and its relationship to defining reality see Neisser [25].

himself. The implication for practice is the recognition that problem definition involves *negotiation* between individuals with different realities, and like all negotiations different interests are represented and different power bases apparent. One of these interest groups is usually the OR department who have survival needs, ideological commitments, and internal politics—these are very much part of the mental schema of the internal consultant and will thus affect his construction of reality. Thus we are suggesting that problem definition must be recognised as a choice activity which involves a form of negotiation between individuals with different interests and power and most particularly that this part of the practice of OR has important consequences for the likelihood of the consultant being successful in doing whatever he wishes to do.

DOUBLE BINDS AND BLUFFS

Stemming from the notion that problems are defined by persons for themselves there is a further area that must be considered; that is that if a person is defining a problem, it will be defined in terms of his world and his experience and what matters to him, and this may include personal factors which are not usually regarded as legitimate topics to mention within the organization [20].

It seems that quite often, we, as consultants, can find ourselves in a bluff position: "I know that you want me to do so-and-so because of such-and-such, and you know that I know, but you are not going to tell me because the 'rules' say you can't". The recognition, by the consultant, of this bluff can be an essential ingredient in a successful consultant–client relationship; good OR consultants often do this implicitly. Indeed it may be that we have to recognise these features of the situation implicitly simply because if we were to do it explicitly then we are displaying to ourselves our involvement in an 'illegitimate' or normatively unacceptable activity. However as we have been seeking to demonstrate there are good reasons why these norms seem to be unhelpful for creating consultant–client interactions that can be satisfying to either party.

This form of analysis of the nature of problems presents a difficult problem for the OR professional which is in the form of a double-bind. If he wants to increase the probability of conducting inquiries that are relevant to the client and result in the client taking action that seems to be a consequence of those inquiries then he will need to pay attention to the unique and particularistic nature of the client's problem. If he does this then he is possibly dissociating himself from what he considers to be the 'needs of the organization' which employs him. If he uses the relationship to further the ends of this entity, which he calls the organization, then he will soon destroy the likelihood of developing the necessary close relationship with clients. Typically this double-bind is ignored rather than consciously and deliberately considered, and this can only lead to a failure to take account of the likely outcomes of paying too much or too little attention to particular interest groups. Alternatively he may maintain a relationship which is designed to solve only the problems that both the client and the consultant regard as organizationally legitimate. In this case either: he is constrained to working on problems where the power to define the problem and force action belongs to him (typical text book OR problems such as inventory control); or he stands little chance of helping the client with the problem the client owns, which thus creates a high probability of his experiencing disillusionment and seeing little action as a consequence of his inquiry.

What does the realization of a double-bind such as the one elaborated above reveal? We consider it means that as a community of OR professionals there is a need to consider and reflect upon the nature of different roles for people with the model building and analytical skills that are the consequence of OR education or experience. Our arguments up to this point lead us to make some assertions about the ways in which it is possible for an operational researcher to work within an organization so that his work has some effect. We see three different ways in which he may attempt to affect the actions of his client:

(i) he may attempt to *coerce* the client into using models and solutions devised by the operational researcher;

(ii) he may attempt to develop *empathy* with the client, discover his definition of the problem, and help him devise a satis-

Vol. 7
1979

factory course of action to dispose of that problem;

(iii) he may behave as we have suggested the client's colleagues behave, and *negotiate* with the client a redefinition of the problem, and subsequently try to help him solve it.

THE COERCIVE PARADIGM

Consultants will often deny the possibility of coercive power for them to influence their client because, they say, the client can ignore their instructions or simply tell them to go away. We maintain, however, that the operational researcher does have a measure of coercive power over his client, and that this power stems mainly from the rational–scientific aura that surrounds the consultant [27]. Operational research may be seen by the client as offering proper, objective, scientific methods for understanding and then solving his problems [15]. He does not expect fully to understand the techniques or the approach, but he is likely to go along with the belief that the application of 'science' and 'reason' must be a good thing [25]. So if the pedlar of science and reason suggests a formulation for a problem and a solution to it, the client who ignores such advice and help may be seen by himself and his colleagues as a backwoodsman, kicking in the face of rationality. That the power of OR techniques appears to practitioners to be limited does not matter for this purpose; if the client's perception inclines him to ascribe this power to the practitioner, he will behave as if this power existed. It is important to consider whether operational research is conducted following a coercive paradigm of the type we have described. If so, we can expect that, for the reasons stated above, it will be unsatisfactory to all parties, leading, as it seems, to the solution of problems which nobody ever had.

THE EMPATHETIC PARADIGM

The consultant may try better to get hold of the problem that has been defined by the client, and then help the client to solve that. We have suggested above some reasons why this may be more difficult than may have been thought. It seems unlikely that consultants knowingly fail to make the effort to grasp the problem defined by their client, but it also seems likely that many operational researchers have tried to grasp that problem while ignoring the organizational factors, as well as the sheer limitations on human communication, which tend to mean that they are given something other than the full problem definition of a client [23]. One of the major disincentives to following this path is that it requires some definite decision about who the client is. We have shown that each person will have their own definition of a problem, but while a group or a team may be able to produce a common wording for a problem definition, there are still likely to be differences in the private definition of that problem held by different members. Thus any consultant who wishes to help with the problem by exploring more fully the definition of that problem held by his client has to decide who he shall regard that client as being, and how he shall relate to persons in the organization whom he does not regard as being his client. If this is made explicit, it is likely to sound, quite realistically, as if he is working on behalf of some person or group and potentially against some other person or group within the organization [13]. As we have noted previously, this may be unacceptable. Perhaps the greater problem with this approach is that it does demand a high degree of understanding by the consultant of the client's problem definition. It may be difficult for the client to make the problem explicit to himself, let alone articulate it to another person. There must be some doubt about the practicality of this approach.

THE NEGOTIATIVE PARADIGM

The consultant may attempt to negotiate, with whomever he has defined to be the client, a redefinition of the problem, in the same sense as the client's colleagues may have negotiated a redefinition with him earlier. Indeed the consultant may conduct this negotiation with the client's colleagues; it is not always easy to discover which member of a work group had the problem in the first place. Indeed by the time the consultant arrives the problem may have been redefined in such a way that there is no longer any one owner. This approach takes more account of the consultant as a person with idiosyncratic beliefs and values, rather

than simply as a reservoir of problem-handling technology. Here it becomes legitimate for the consultant to seek a redefinition of the problem in such a way that it makes sense to himself. To do this he will need to be self-conscious about where his expertise does and does not lie, to be careful in his use of that knowledge, and particularly to be careful that he does not speak and behave in an expert manner when outside his field of expertise; otherwise he is still exercising coercive power.

Let us consider more formally the most common situation for an OR worker. An organization employs a group of operational researchers who co-exist within an OR department. Their survival can never be totally assured for service departments are seen as the most easily dispensable. Their clients are likely to be a manager or a small group such as a project team. This situation suggests that the client owns some sort of problem which has been defined within the context of his world—his definitions of organizational needs, his political environment, his own needs and aspirations. The consultant is ready to prosecute a role that reflects his own definition of organizational needs, his political environment—in particular his departmental setting, his own needs and ethics. Neither of these entry points are *known* to the other, nevertheless assumptions must be made and images constructed in advance. At the beginning of an interaction the client has an image of the role of the consultant and a set of what he may regard as realistic expectations [8]; these may be derived from past experience of OR consultants, from the journals, from the library bookshelf contents or from a knowledge of the education and experience of the consultant. The point is that there will be a set of expectations constructed that will affect the interaction because they are the self imposed rules that will be used for deciding what is legitimate and appropriate to discuss with the consultant. For example the client 'knows' that OR is to do with mathematical modelling, he knows that the last piece of work that the consultant did was for the fellow who runs the other division, he knows that the OR department is a part of head office—it is easy to see how this will impact on the selective presentation of the 'problem' to the consultant.

The problem will, of course, have features which may be represented mathematically; it is likely to involve inquiry that will reveal to the client something of the nature of the problem the other division declared; it will not have features that could reflect badly upon the operation of the manager's domain in head office terms, probably the opposite; and yet it will all be made to seem to make good sense to the consultant as a potentially important problem worth solving—for it is also likely that the client believes the consultant wants a problem that he can solve and make recommendations about. For the consultant's part he has the context of the reality of the organization he has formed from working on other studies and he has a need to do something which can be acknowledged publicly by his colleagues and head office employers. The presentation by the client sounds like a legitimate 'public' problem which enables him to use those well structured skills contained by a systems frame of reference—the art of establishing system objectives, system input and output, the harmonious activities of "object system and subject system" [29]—the model may not recognise the realities of organizational life but then the client never stated these! The consultant has legitimised his going off and inquiring, building a mathematical model, solving, and reporting recommendations. Clearly this is one course of action for the crude survival of the OR person. It is also clear that many potential gains are not realised by any of the interested parties (as Brewer demonstrated in his study), although many competent operational researchers cope with these subtle complications intuitively. Our own enthusiasm is directed to making success within the profession less dependent upon intuition. Often intuition only enables the consultant to protect his rear, while all parties remain disillusioned with the consequences of his energy. Additionally the on-going enactment of this type of OR consultation reinforces the set of expectations clients have of the type of work OR people can and will do.

THE KING IS DEAD, LONG LIVE THE KING?

What we have said above has consequences for the practice of OR which are not easy to handle, but which are significant and should

Vol. 7
1979

not be avoided. We shall argue in the following sections that the contract which is made between an OR consultant and his client needs to reflect more realistic expectations. We shall also be arguing that the nature of problems and problem definition, as it has been outlined above, has implications for the relationship between the OR practitioner and his client as well as for the area to which that practitioner should be paying attention in his practice. We consider these consequences to be difficult to handle because each one of them has some major implications for the way in which practitioners go about the business of doing OR. The difficulty of handling these consequences is compounded by the fact that they are in terms of dimensions of thinking to which attention is not normally paid by operational researchers. In other words, *we are not merely redirecting attention from one set of dimensions to another, but we are suggesting that extra (and for the operational researcher, unfamiliar) dimensions need to be taken into account, and such a suggestion must necessarily increase the complexity of OR.*

THE ROLE OF THE CONSULTANT

The first area of consequences is for the kind of function which OR can be expected to fulfil. We see two major contenders for such a role; one is a kind of technician role, in which somebody else must be responsible for formulating the problem, for deciding which criteria of evaluation are significant, and the operational researcher is left with the role of providing the techniques to find and run data to fit these requirements from a repertoire of techniques which he carries with him.

If the problems are complex ones, the communication between the person who has the problem and the operational researcher is complicated enough to make it difficult for the problem owner to be sure that the operational researcher is solving the 'right' problem, or that the way he is solving it is really taking into consideration the factors that the problem owner wanted to be taken into consideration. If there is a doubt in the problem owner's mind about whether the correct factors have been taken into account, he is unlikely to believe the solution to the problem, and if he does not believe the solution he will not act upon it.

The alternative function or role for the operational researcher takes him away from being involved in problem solution into a more general help role. In this role it will be his function to assist or enable his client in working with his problem, using all appropriate abilities and techniques at his command. However the practitioner must be prepared to negotiate from the client's version of what he believes his problem to be, and the relationship which he believes to exist between the elements of the problem. In a help role, the practitioner will need to resist the temptation to collect data or to apply techniques which he would like to do, without the committed consent of the client. The operational researcher will have to forgo the luxury of claiming a preemptive expertise in the solving of problems. Against this we would argue that by taking a role of collaborative help towards the client, he is more likely to produce credible solutions to problems which the client actually has.[5]

DISCOVERING A PROBLEM

A significant early stage in the intervention process is that of developing an acceptable degree of empathy for the client. That is the consultant must become aware of the world as it looks from the standpoint of the client. An attempt to formally map aspects of the defined problem in the terms of concepts used by the client can be a successful introduction to the process of problem formulation.

As we have said earlier, in many cases the client is either a group of persons or the initial problem definer has had to take his problem through a group before it has reached the consultant. This implies that the consultant will need skills for facilitating the negotiation of a substantially shared problem which reflects the power and interests of the members of a group. When a problem is stated as being that of a group of people, and is presented by a group

[5] Clark [14] offers a very helpful summary of the different role types operated by consultants in the context of action research, see Chapter 7.

of people, two things must be considered. Firstly, the understanding of the problem becomes more complicated as the consultant has to take more perspectives into account. Secondly, the political situation underlying that problem and forming the background to the activities of the group at that time are of critical importance to the consulting activity. The consultant who fails to use group processes within the organization in which he is going to assist with data collection, is almost certainly depriving himself of his best but also his most complicated data. In addition the group is the most likely arena for whatever actions are instituted in order to deal with the problem.

The third area of consequences has to do with the type of conceptual skills that the consultant needs for his work. Most forms of training of OR practitioners impart a body of skills which is concerned with the construction and manipulation of explicit models of the problem, and usually of technical and readily quantifiable aspects of the problem. From our discussion above of the nature of problems we would suggest that, as well as more sophistication in his choice of role and in the process of understanding the problem owned by the client the OR consultant could also play a less restricted part if he could construct his models to take account of more psychological and sociological factors. It is difficult to conceive of a consultant modelling for himself as simple a set of criteria, values and beliefs as those which he normally ascribes to a client. It seems likely that the operational researcher who wishes to become involved in complex issues of social and technical significance would make his work both more difficult and more effective if he had an anthropomorphic model of his client, and assumed his client to be a confused, semi-rational being like himself, with uncertain values and beliefs, and uncertain relationships among these values and beliefs.

Additionally if the consultant does not pay explicit attention to power, or influence, or interpersonal processes in groups, or intra-personal confusion, or conflict and competition between persons, then he will find himself, at least in the eyes of many clients, to have little to say and not much help to offer on those issues and problems which are the very stuff of organizational life [1].

SOME BASES FOR DEVELOPING PROBLEM FORMULATION TECHNIQUES

Our argument so far cannot lead us to any unique solution to the issues and problems discussed [9]. Nevertheless we, like others in the field, have been actively searching for and experimenting with, techniques with which to make useful and relevant contributions to a client's handling of the problems which he thinks are important.

For some time we have been successfully developing and using a set of techniques for getting to know about the significant aspects of the problem that is constructed by the client or client group. These techniques are, for the most part, centred on the outcome of an ongoing interaction between consultant and client, where the consultant is seeking to make sense of and explore the ramifications of the qualitative and quantitative experiences and feelings the client has about his problem. They have been developed by combining a range of apparently disparate techniques. We have used 'construct grids' derived from clinical psychology [6] to understand more about the complex system of qualitative objectives that underlie ill-structured problems [3, 21]. This method of collecting data about preferred outcomes combined with methods for structuring and analysing the beliefs the decision-maker has about the causes and consequences of his problem usually lead to the explicit representation, or map, of a formulated/structured and owned problem. This 'cognitive map' is the starting point for analysis, the development of an inquiry strategy, and the exploration of possible solutions—all of which can be undertaken with the assurance of commitment from the client. Cognitive mapping, which originates in the theory of directed graphs and has been exploited by political scientists [4], can also be used successfully in combination with Nominal Group Techniques and the Delphi technique [18] for facilitating the negotiation of an 'owned' problem when the client is a small group [22]. Further analysis opportunities can often be developed by the conversion of cognitive maps into the causal maps used as the basis for System Dynamics simulation modelling [16].

Vol. 7
1979

CLOSING REMARKS

In our introduction we suggested that even though OR practice was often highly successful and development of "a more sound theoretical basis for understanding and reflecting upon the *process* of effective consulting" would enable many consultants who are currently anxious about their intervention style to become better and less disillusioned practitioners. We have attempted to illustrate some of the critical features we believe should be contained within a theoretical framework or heuristic which can guide the continuous and reflective development of *theories of practice*. We are therefore concluding with a plea that the OR community allow more time and energy for reflection in order to develop what Argyris and Schon call 'Theories in Action' about professional practice [2].

For this to occur a high level of commitment to seeing this reflection as an everyday activity is necessary, for it is time-consuming to develop and test out theories-in-action. In addition because we are not used to operating a new theory and understanding events with it, it will be more difficult and energy consuming to operate this theory than to operate one to which we are more habituated, because we will have to put conscious thought in to considering the meaning, relationship and consequences of events under that theory. This time and energy is unlikely to show any immediate consequences visible to the practitioner's colleagues or superordinates, and there is not likely to be any immediate incentive for an OR consultant to spend time this way unless he makes a deliberate decision to do so. If, however, the OR consultant wishes to be an effective helper to his client, it seems essential that he should reserve time and energy for continuous development and improvement of his own theories and models about the organizational as well as the technical contexts of his clients, and about himself as a person and as a theorist within those contexts.

REFERENCES

1. ACKOFF RL (1978) The future of operational research is past. Paper presented to the Opl Res. Soc. Conf., University of York. *J. Opl. Res. Soc.* **30** (2), 93–104.
2. ARGYRIS C & SCHON DA (1974) *Theory in practice: Increasing Professional Effectiveness.* Jossey-Bass, San Francisco, USA.
3. ARMSTRONG & EDEN (1978) An exploration of occupational role. *Personn. Rev.* Winter.
4. AXELROD R (1976) *Structure of Decision.* Princeton University Press, New Jersey.
5. BALL DW (1972) The definition of the situation: some theoretical and methodological consequences of taking WI Thomas seriously. *Jl Theory Soc. Behav.* **2**(1), 61–82.
6. BANNISTER D & FRANSELLA F (1971) *Inquiring Man.* Penguin, Harmondsworth, Mddx.
7. BITTNER E (1973) Objectivity and realism in sociology—*Phenomenological Sociology: Issues and Applications* (ed Psathas F). John Wiley, New York.
8. BOOTHROYD H (1977) Basic thoughts for sovereign minds—*The King is Dead: Long Live the King.* Proc. of one-day Conf. Opl. Res. Soc. Birmingham.
9. BOWEN KC (1975) Where techniques are lacking. *Def. Opl. Analysis Establ.* West Byfleet.
10. BOWEN KC (1977) The process of OR—a systems approach—*The King is Dead: Long Live the King.* Proc. of one-day Conf. Opl. Res. Soc. Birmingham.
11. BREWER GD (1973) *Politicians, Bureaucrats and the Consultant.* Basic Books, New York.
12. BRYER RA (1977) On the status of the systems approach—*The King is Dead: Long Live the King.* Proc. of one-day Conf. Opl. Res. Soc. Birmingham.
13. BURNS T (1965) On the plurality of social systems—*OR and the Social Sciences* (ed Lawrence JR). Tavistock, London.
14. CLARK PA (1972) *Action Research and Organizational Change.* Harper & Row, London.
15. COLLINS R (1975) *Conflict Sociology: Toward an Explanatory Science.* Academic Press, New York.
16. COYLE RG (1977) *Management System Dynamics.* Wiley, London.
17. DANDO M & EDEN C (1977) Reflections on operations research: a report from the EURO II Congress. *Omega* **5**(3), 255–270.
18. DELBECQ AL, VAN DEN VEN AH & GUSTAFSON DH (1975) *Group Techniques for Program Planning.* Scott, Foresman & Co, Glenview, Illinois.
19. EDEN C (1975) An exploration of the role of a decision analyst. University of Bath.
20. EDEN C (1977) Modelling the influence of decision makers on the future. *Futures* **9**(4) 272–284.
21. EDEN C (1978) Operational research and organization development. *Hum. Relat.*, **31**(8), 657–674.
22. EDEN C & JONES S (1978) Subjectivity—a case study presented to the *OR Soc. Conf.* University of York.
23. EDEN C & SIMS D (1976) Problem definition between consultant and client. University of Bath.
24. MCHUGH (1968) *Defining the Situation.* Bobbs-Merrill, New York.
25. NEISSER U (1976) *Cognition and Reality.* WH Freeman, San Francisco, USA.
26. REIN M & SCHON DA (1977) Problem setting in policy research—*Using Social Research in Public Policy Making* (ed Weiss CH). Lexington Books, Lexington, Mass. USA.
27. SANDBERG A (1976) The role of rational analysis in decision-making—harmony or conflict perspective? Presented to EURO II, Stockholm.
28. SIMS DBP (1978) A framework for understanding the definition and formulation of problems in teams. Paper presented to *Brit. Assoc. Advmt Sci. Annual Meeting*, University of Bath.
29. WHITE DJ (1975) *Decision Methodology.* John Wiley, New York.

ADDRESS FOR CORRESPONDENCE: *Dr Colin Eden and Dr David Sims, School of Management, University of Bath, Claverton Downs, Bath BA2 7AY.*

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 2. pp. 129 to 135
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0401-0129$02.00/0

The Disclosure of Information by Companies

MICHAEL FIRTH

Victoria University, New Zealand

(*Received July* 1978)

The paper presents some results relating to the degree of disclosure of selected important items of information in company annual reports. The research involved deriving an index of disclosure and this was then applied to the annual reports of 100 stock exchange quoted companies. The general conclusion was that disclosure levels are very low and that there is a big demand for greater amounts of information to be released in annual reports.

ol. 7
979

THE DISCLOSURE OF INFORMATION BY COMPANIES

DURING THE 1970's there has been a growing interest and debate in the area of corporate financial reporting in the UK. For example the professional accounting bodies commissioned a study into the whole nature of financial reporting and this resulted in the publication of The Corporate Report [3]. Other studies have, for example, looked at the degree of understanding of annual reports by private shareholders [19], the degree of consensus about the usefulness of accounting data [17], and the impact of accounting information on share prices [15, 16]. The purpose of the present paper is to add to the literature on financial reporting by examining the degree of disclosure actually achieved by British stock-exchange quoted firms.

The research involved constructing a list of individual items that are, or could be, included in the annual report. Financial analysts working for stockbrokers and investment institutions were then asked to weight these items in terms of importance. A set of 100 stock exchange quoted companies was selected and their published annual reports were examined to see whether the items appeared. From this examination the extent of disclosure of each of the individual items was ascertained and the relationship between the importance of an item and the extent of its disclosure was measured.

Similar studies, using disclosure indexes, have been undertaken in the USA, [9, 10, 11, 12, 23] and recommendations have been made as to how to improve the utility of the annual report data. For example, Buzby found that segmental reporting of sales turnover and of income were deemed to be very important by financial analysts and that, in practice, very little such disclosure is being made by companies [9].

RESEARCH METHODOLOGY

List of items

The first step in the research design was to construct a list of items of information that appear, or could appear, in corporate annual reports. A review of the relevant literature, [5, 6, 8, 9, 10, 12, 13, 14, 18, 21, 23] was made in order to select the items to be included in the list. Although this is not an exhaustive list it does contain those items most frequently mentioned in the accounting press as being useful for investors and bankers. A review of several recent company annual reports and discussions with various users of accounts was also used in order to refine the list. Finally the list was subjected to two criteria; the first is that the items should exclude those that had to appear

in annual accounts because of statutory regulations (e.g. Companies Acts)[1]. The reason for this criterion is that companies would almost certainly disclose this information. The second criterion is that the item was almost certainly bound to be present in a company. For example manufacturing firms (which make up the sample) engage in research and new product development and they can therefore publish a narrative summary covering this, if they so wish. Any companies which definitely did not have anything to do with the items were omitted.

Importance of items

Although we now have 48 items of information it is quite likely that some are of more importance to the users of accounts, than others. In order to take this into consideration 'importance' weightings were attached to the items. This was done by sending the list of items to users of accounts and asking them to evaluate the importance of each item using a five point scale (see Appendix). Assigning the score of 5 meant the item was very important, the score of 4 meant it was important, the score of 3 related to moderately important items, the score of 2 related to slightly important, and the score 1 meant unimportant.

The list was sent to a total of 120 financial analysts working for stockbrokers and investment institutions. Forty-six replies were received which represents a response rate of 38.3%. Although the response rate is quite reasonable when compared against other studies it is possible that there is some non-response bias (i.e. that the results from those responding may not be truly representative of the whole population). To examine whether there was any significant non-response bias present, a test suggested by Oppenheim [20] was employed. This involved examining whether the mean scores for each of the 48 items for the first ten questionnaires returned were statistically different from the mean scores for each of the 48 items for the last ten questionnaires returned. The results of the *t*-test, which examined for differences in group means, are shown in Table 1. As can be seen there were few significant differences and the signs of the *t* values were equally split between positive and negative. Hence there is little significant non-response bias and this suggests that the results of the study can be generalized.

TABLE 1. NON-RESPONSE BIAS: RESULTS OF *t* TEST

Number of items	48
Number of significant differences in mean scores*	2
Positive *t* statistics	24
Negative *t* statistics	24

* Where the *t* statistic was significant at the 0.05 level.

The questionnaire containing the 48 items was sent to financial analysts as they are one of the major types of user of accounts and they are in a good position to evaluate the importance of the items. Previous research has shown that the weightings applied by financial analysts were very similar to those applied by bank loan officers, who are another major type of user of accounts [17][2]. Thus the results of the study would be similar even if the weights of bank loan officers were used in evaluating the importance of the items.

A list of the 48 items and the weightings attached to them is shown in Table 2; the list has been prepared in the order of relative importance. The higher the score the greater the importance that that item should appear in the Annual Report. The 'breakdown of sales and earnings by product lines, customer classes and geographical location' proved to be the most important item to be disclosed and the 'historical summary of price range of ordinary shares in past few years' proved to be the least important.

As a general rule, historical accounting data tended to receive fairly high importance weightings. Forecasts of the future (e.g. forecast of next year's profits and cash projections for the next 1–5 years) received 'moderate' to 'important' scores—this suggests that financial analysts were placing some doubt on the accur-

[1] A few items (e.g. numbers 1 and 2 in Table 2) were covered by the Companies Acts but they were included as the wording in the Acts is 'loose' (e.g. as respects materiality) and many firms do not comply. Additionally some items are included which are covered by Statements of Standard Accounting Practice (S.S.A.P.'s), published by the Accounting Standards Committee, and are therefore supposed to be reported on by companies. As some companies do not follow certain S.S.A.P.'s and as many S.S.A.P.'s are considered to be important by investors, these items have been included.

[2] This contrasts with the USA, however, where Benjamin and Stanga [7] found that there were significant differences between chartered financial analysts and bank loan officers in their weightings of selected accounting items.

Vol. 7
1979

acies of such forecasts. However, the mean scores and relative rankings of forecasts were higher than those found in the USA, [7, 9, 10, 13].

The Corporate Report [3] recommended that a number of additional statements should be included in the annual report. These statements were included in the list of items (see items 18, 24, 30, 40, 41, 44, 45) and quite high scores were recorded for 'statement of company objectives', 'statement of value added' and 'a statement of transactions in foreign currency'. Although few such statements have appeared in practice, financial analysts appear to be very interested in seeing such information.

The provision of inflation-adjusted annual accounts as supplementary statements (as suggested by E.D.18, [4]) received a fairly high score. This contrasts with the USA where Benjamin and Stanga [7] and Chandra [13] found that very low weightings were attached to inflation adjusted accounts.

Sample firms

The disclosure index, made up of the 48 items and their weightings, was then applied to a sample of the actual annual accounts of stock exchange quoted companies. The sample consisted of 100 manufacturing companies—these being the every 10th firm appearing in *The Times* 1000 largest companies [24]. If the tenth firm was not a manufacturing firm, then the next manufacturing firm was used. The size of the companies in the sample ranged from over £3,500 million in terms of capital employed to just under £1 million. Whilst these are the largest companies they are also the ones which most interest investors and hence adequate disclosure is perhaps more important for them. Annual reports for the 100 companies, released during 1976, were obtained and used as the data in the study.

RESULTS

The next stage of the analysis was to determine the extent of disclosing the items by the 100 companies. As explained earlier the items were thought to be present in every company and the study sets out to see whether the information is disclosed. The results of the analysis are set out in Table 2. Column 2 gives the weightings attached by financial analysts, and column 3 ranks these in descending order of importance. The number of firms receiving a zero for disclosure (i.e. they did not disclose the specific item in the annual report) is shown in column 4, and column 5 shows the number of companies who received the full score for disclosure (i.e. they did disclose the specific item). The average disclosure score for each item is given in column 6. Column 7 expresses the average disclosure score as a percentage of the maximum possible score; if this came out as 0 then it means the item was not being disclosed at all, if the figure came out as 1.00 then it means all firms in the sample were disclosing the item. Column 8 gives the ranking in terms of disclosure i.e. item 14, 'discussion of the company's results for the past year' was the one which received the most disclosure in the sample.

In general the results indicate that disclosure levels in the UK are very low. Only 8 items out of 48 were reported on by more than half the sample population, and 8 items were not reported on at all (this including some items, e.g. cost of sales, that were considered to be extremely important). On average only 19.75% of companies were reporting the individual items. It would appear that British companies do little more than disclose information that has to be reported because of legislation, principally the Companies Acts of 1948 and 1967, and the Statements of Standard Accounting Practice published by the Accounting Standards Committee. The force of opinion for disclosing some items of information (e.g. items 18, 24, 30, 40, 41, 44, 45) has only recently been felt and hence companies may still be in the throes of formulating exactly how these items should be reported. Apart from this the major reasons for the lack of disclosure are likely to be, firstly, that finance directors, and to some extent their auditors, who are responsible for preparing accounts, may not be aware of the importance of some items for the users of accounts. This is borne out by a recent research study in the UK which showed that finance directors and auditors have greatly differing views from financial analysts and bank loan officers on the importance of information[3]

[3] The study analyzed 75 items of data including the 48 used in the current study.

TABLE 2. DISCLOSURE INDEX

Item number	Item	2 Financial analysts' weight	3 Rank	4 Number of firms with zeros	5 Number of firms with maximum score	6 Average score	7 Average as a percentage of column 2	8 Rank
1	Breakdown of sales revenue by major product lines, customer classes and geographical location.	4.82	1.5	37	63	3.04	0.63	5
2	Breakdown of earnings by major product lines, customer classes and geographical location.	4.82	1.5	39	61	2.94	0.61	6
3	Cost of goods sold.	4.71	3	100	0	0.00	0.00	44.5
4	Amount and detailed breakdown of expenses (beyond that required by the Companies Acts).	4.69	4	95	5	0.23	0.05	26
5	Breakdown of expenses for past year into fixed and variable components.	4.63	5	100	0	0.00	0.00	44.5
6	Statements of source and application of funds.	4.58	6	9	91	4.16	0.91	3
7	Information relating to capital expenditures (narrative and quantitative data on expenditure in past year and planned expenditure).	4.51	7	81	19	0.86	0.19	13
8	Information relating to research and development (including a narrative description of the progress with new product development, planned expenditures).	4.46	8.5	83	17	0.76	0.17	14
9	Share of market in major product areas.	4.46	8.5	95	5	0.22	0.05	26
10	Names and salaries of senior management.	4.42	10	96	4	0.18	0.04	30.5
11	Index of quantity sales.	4.41	11	100	0	0.00	0.00	44.5
12	Statement of accounting principles used in the preparation of the accounts.	4.40	12	7	93	4.09	0.93	2
13	Historical summary of important operating and financial data.	4.39	13	13	87	3.82	0.87	4
14	Discussion of the company's results for the past year.	4.27	14	3	97	4.14	0.97	1
15	Measures of the physical level of output and capacity utilization.	4.26	15	92	8	0.34	0.08	22.5
16	Index of sales prices.	4.25	16	100	0	0.00	0.00	44.5
17	Functional responsibilities of senior management.	4.24	17	96	4	0.17	0.04	30.5
18	Statement of company objectives.	4.21	18.5	89	11	0.46	0.11	18.5
19	Specification of the method used to calculate depreciation (including depreciation rates).	4.21	18.5	48	52	2.19	0.52	7
20	Index of raw material prices.	4.20	20.5	100	0	0.00	0.00	44.5

21	Money value of firm's order backlog.	4.20	20.5	91	9	0.38	0.09	21
22	Inflation adjusted annual accounts as supplementary statements.	4.16	22	80	20	0.83	0.20	11.5
23	Extent of dependence on a few customers.	4.08	23	95	5	0.20	0.05	26
24	Statement of value added.	4.04	24	95	5	0.20	0.05	26
25	Functional responsibilities of company directors and their major outside affiliations.	4.01	27	92	8	0.32	0.08	22.5
26	Statement of the rate of return required by the company on its projects.	4.01	27	100	0	0.00	0.00	44.5
27	Description of major products produced.	4.01	27	85	15	0.60	0.15	15
28	Information relating to advertising and publicity (e.g. expenditures in past year, future commitments).	4.01	27	88	12	0.48	0.12	17
29	Market value of inventory.	4.01	27	97	3	0.12	0.03	34
30	A statement of transactions in foreign currency.	3.99	30	99	1	0.04	0.01	39
31	Allowance for doubtful debts.	3.98	31	98	2	0.08	0.02	36.5
32	Discussion of the major factors which will influence next year's results including an indication of the firm's relationship to its industry and the economy.	3.96	32	71	29	1.15	0.29	9
33	Brief narrative history of the company.	3.95	33	90	10	0.39	0.10	20
34	Discussion of the impact of inflation on the financial results.	3.92	34	49	51	2.00	0.51	8
35	Information on major industry trends.	3.89	35	80	20	0.78	0.20	11.5
36	Cash projections for the next one to five years.	3.87	36.5	95	5	0.19	0.05	26
37	Current resale value of the firm's assets.	3.87	36.5	98	2	0.08	0.02	36.5
38	Number and type of employees.	3.69	38	97	3	0.11	0.03	34
39	Forecast of next year's profits.	3.67	39	72	28	1.03	0.28	10
40	Indication of employee morale (i.e. labour turnover, strikes, absenteeism).	3.55	40	99	1	0.04	0.01	39
41	Information on corporate social responsibility (i.e. attitude of the company, expenditures)	3.48	41	89	11	0.38	0.11	18.5
42	Summary of the age of debtors at the balance sheet date.	3.27	42.5	100	0	0.00	0.00	44.5
43	Description of major plants, warehouses and properties, including location, function and size.	3.27	42.5	97	3	0.10	0.03	34
44	A statement of money exchanges with the Government.	3.21	44	99	1	0.03	0.01	39
45	Amount expended on human resources (training, welfare facilities).	3.01	45	87	13	0.39	0.13	16
46	Number and type of ordinary shareholders (e.g. institutions, individuals).	2.51	46	96	4	0.10	0.04	30.5
47	Description of marketing network for finished goods.	2.48	47	100	0	0.00	0.00	44.5
48	Historical summary of price range of ordinary shares in past few years.	1.79	48	96	4	0.07	0.04	30.5

[17]. This is unlikely to be the whole reason, for even if we substitute the finance directors weightings (see [17]) for the financial analysts, there is still a substantial shortfall in disclosure. This leads us to a second reason, namely that finance directors feel that the disclosing of information, such as cost of sales and share of market,would be giving important data to their competitors which is not being reciprocated. The giving away of valuable data to competitors is a valid enough argument and therefore legal or quasi-legal (perhaps via the Accounting Standards Committee) regulations may be needed to make greater disclosure compulsory for all. Of course such regulations would only apply in the UK and hence companies may be at a disadvantage with their international competitors who do not have extensive disclosure rules.

The results are similar to those found in the USA by Buzby [9]. Buzby did in fact find large disclosure scores for some items but these tended to be those that have to be reported in the UK because of the Companies Acts regulations. For the 48 items included in the current study, Buzby also found very low disclosure levels.

Although the results are fairly self-explanatory a number of features can be discerned. Firstly, items of information that are recommended to be shown by the professional accounting bodies, e.g. Statements of Source and Application of Funds [2], and Statement of Accounting Principles [1], receive high disclosure levels.

So far, few firms appear to have incorporated inflation adjusted accounts in their annual reports even though this is considered to be quite important information by financial analysts. This disclosure shortfall may largely be due to the problems that have beset the Accounting Standards Committee in trying to establish a feasible and satisfactory method of inflation accounting. In due course, no doubt, inflation adjusted accounts appearing (at least) as supplementary statements will become the norm. This is evidenced by the willingness of management to refer to inflation when discussing the performance and prospects of the company (see item 34).

Items that relate to supplementary statements recommended by The Corporate Report tended to receive lowish scores even though many of them do not contain information that is likely to be valuable to competitors. This lack of disclosure is likely to be the result of finance directors both thinking that the items are not important and that they have not had enough time to consider how to report such information.

Items 32 and 29 relating to forecasts of the future received relatively high disclosure scores, although item 36, cash projections for the next one to five years, had a low score. The results contrast with the USA where forecasted data received both a low importance weighting and a low disclosure score. The fact that forecasted information is being given by many companies in the UK may suggest that the regulatory bodies, such as the accountancy profession and the Stock Exchange, should consider giving some guidelines as to their form and content.

An important class of items are those relating to the breakdown of expenses (items 3, 4 and 5). These received extremely low disclosure scores, almost certainly because of the 'giving away of information to competitors' reaction. Because of this lack of disclosure, financial analysts and others will only be able to guess at the cost structure and cost-volume-profit profile of companies.

An analysis was also made to see if an item's perceived importance weighting was associated with its disclosure level. This was undertaken as it can be argued that companies only disclose information if it is important and they omit data which is not of much value to the users of accounts. Spearman's rank correlation test [22] was used to measure the association between the rankings of the importance weightings (column 3) and the rankings of disclosure levels (column 8). The correlation coefficient was 0.24. This indicates that there is little association between the importance of an item and the extent of its disclosure (similar conclusions were reached even when the finance directors' weightings were substituted for the financial analysts' weightings in the research design).

CONCLUSION

The object of the research was to measure the extent of disclosure achieved by stock exchange quoted British companies in their annual reports. It was found that disclosure

Vol. 7
1979

levels were very low with many items of information not being referred to at all. The vast majority of companies appear to report little more than what they have to under the Companies Acts. Two major reasons for the shortfall in disclosure were postulated, these being that financial directors were unaware of the importance of some items and hence they did not report, and to prevent competitors gaining access to, possibly important, operating data. The position as regards the first reason could be improved by greater publicity being given to the differences in the perceived importance of items between the preparers of accounts and the users of accounts—this could lead to greater and more relevant information being released by companies. The second reason may need legal or quasi-legal rulings to secure greater disclosure so that all firms, in Britain at least, are on an equal footing. A further analysis showed that there was little relationship between the importance of an item and the extent of its disclosure—thus the hypothesis that companies only bother to disclose information which is deemed to be very important is not upheld. The general conclusion from the research is that disclosure levels are very low and, that while there may be some demand for more simplified accounts, there is also a big demand by some users of accounts for greater amounts of information to be released in company annual reports.

REFERENCES

1. Accounting Standards Steering Committee (1971) Disclosure of accounting policies. *Statem. Stand. Acctg. Pract.* **2**.
2. Accounting Standards Steering Committee (1975) Statements of source and application of funds. *Statem. Stand. Acctg. Pract.* **10**.
3. Accounting Standards Steering Committee (1975) The corporate report. *Acctg. Stands Steerg. Comm.*
4. Accounting Standards Committee (1976) *Current cost accounting.* Exposure Draft 18.
5. Anderson CD (1962) *Corporate Reporting for the Professional Investor.* Fin. Analysts Fed.
6. Backer M (1970) *Financial Reporting for Security Investment and Credit Decisions.* Nat. Assoc. Accnts N.Y.
7. Benjamin JJ & Stanga KG (1977) Differences in disclosure needs of major users of financial statements. *Acctg Busin. Res.* **27**, 187–192.
8. Bradish RD (1965) Corporate reporting and the financial analyst. *Acctg Rev.* **40**(4), 757–766.
9. Buzby SL (1974) Selected items of information and their disclosures in annual reports. *Acctg Rev.* **49**(3), 423–435.
10. Buzby SL (1975) Company size, listed versus unlisted stocks, and the extent of financial disclosure. *Jl Acctg Res.* **13**(1), 16–37.
11. Carpenter CG, Francia AJ & Strawser RH (1971) Perceptions of financial reporting practices: an empirical study. *MSU Busin. Topics* **19**(4), 56–62.
12. Cerf AR (1961) *Corporate Reporting and Investment Decisions.* University of California, Berkeley.
13. Chandra G (1974) A study of the consensus on disclosure among public accountants and security analysts. *Acctg. Rev.* **49**(4), 733–742.
14. Choi FDS (1973) Financial disclosure and entry to the European capital market. *Jl Acctg Res.* **11**(2), 159–175.
15. Firth MA (1976) The impact of earnings announcements on the share price behaviour of similar type firms. *Econ. Jl.* **86**, 296–306.
16. Firth MA (1978) Qualified audit reports: their impact on investment decisions. *Acctg Rev.* **53**(3), 642–650.
17. Firth MA (1979) A study of the consensus of the perceived importance of disclosure of individual items in corporate annual reports. To be publd. *Int. Jl. Acctg. Ed. & Res.*
18. Horngren CT (1970) Implications for accountants of the uses of financial statements by security analysts. *Compustat Inf. Man. Stand. Statist. Co.*
19. Lee TA & Tweedie DP (1977) *The Private Shareholder and the Corporate Report.* Inst. Chart. Acctnts. Engl. Wales.
20. Oppenheim AN (1966) *Questionnaire Design and Attitude Measurement.* Heinemann, London.
21. Pankoff LD & Virgil RL (1970) On the usefulness of financial statement information: a suggested research approach. *Acctg Rev.* **45**(2), 269–279.
22. Siegel S (1956) *Nonparametric Statistics for the Behavioural Sciences.* McGraw Hill. New York.
23. Singhvi SS & Desai HB (1971) An empirical analysis of the quality of corporate financial disclosure. *Acctg. Rev.* **46**(1), 129–138.
24. Times Newspapers (1977) *The Times* 1,000, 1977–1978. Times Books.

ADDRESS FOR CORRESPONDENCE: *Michael Firth, Department of Accountancy, Victoria University, Private Bag, Wellington, New Zealand.*

APPENDIX

The enclosed questionnaire lists items which could be presented in the annual report of a manufacturing company. You are requested to examine each item and attach a weight to it reflecting how important you think it is that it appears in corporate annual reports. You should consider yourself as making a decision as to whether to invest in a firm. The weights to be used are

5 where the item is very important
4 where the item is important
3 where the item is moderately important
2 where the item is slightly important
1 where the item is unimportant

Please place your view of the importance weighting in the column headed 'weighting' opposite each item. Consider the firm concerned as being quoted on the United Stock Exchange and that each item is 'material' or significant.

In assigning the weights remember that the reporting of each additional item of information involves a cost to the company. Please note that there is no requirement that you should assign an equal number of weights.

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 2. pp. 137 to 143
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0401-0137$02.00/0

An Hierarchal Location–Allocation Problem

SUBHASH C NARULA

Rensselaer Polytechnic Institute, New York, USA

UGONNAYA I OGBU

Ugogbu International, Nigeria

(*Received April* 1978; *in revised form November* 1978)

We consider a 2-hierarchal location–allocation problem when p_1 health centers and p_2 hospitals are to be located among n potential locations so as to minimize the total weighted travel distance. We consider the possibility of the referral of θ $(0 \leq \theta \leq 1)$ proportion of patients from a health center to a hospital. For a graph with certain properties, we extend the notion of a median of a graph to an hierarchal median set. We show how the 2-hierarchal location–allocation problem may be solved as a 2-hierarchal vertex median set problem. We formulate the problem as a mathematical programming problem, suggest five heuristic procedures to solve the problem and report some computational experience. Although we consider the hierarchal location–allocation problem with reference to a health care delivery system, the results of this paper have wide application.

1. INTRODUCTION

CONSIDER health care delivery systems in developing nations. As most developing nations do not have enough financial resources and/or trained medical personnel to build and staff adequate health care facility in every town, hierarchal system of health care facilities has been recommended by several authors, e.g. Bryant [1], Fendall [4] and Gish [5]. An hierarchal health care delivery system may consist of health centers, hospitals, and medical centers. Such facility systems are generally successively inclusive. That is, in such a system, health centers provide primary care; hospitals provide primary care and some additional services; and medical centers provide specialized care in addition to services available at a hospital and a health center. Thus a patient who is not completely cured at a health center may be referred to a hospital or a medical center for treatment; and a patient who first goes to a hospital may be referred to a medical center, if necessary. A medical center is equipped to treat all cases. Navarro [11] reported a successful implementation of such a system in Cuba.

For a health care delivery system that consists of health centers, hospitals and medical centers, Schultz [13] gave a procedure to determine the optimal service radius for each type of facility. This is equivalent to finding the distance at which the per capita net social benefit and the *per capita* cost are equal to each other. For a homogeneous uniformly populated region, he used the central place theory to show that a repeating congruent arrangement of facilities is the optimal pattern. Schultz' model is not a realistic representation of a practical location–allocation problem because his assumption of homogeneous uniform population is not valid in practice. Moreover, he does not consider referrals of the patients from a lower to a higher level facilities. The number of facilities of each type are also not specified in advance.

For the same type of health care delivery system, Calvo and Marks [2] did not assume a homogeneous uniform population for the region. However, they assumed that the

patients at each population center can be divided into $k(=3)$ groups according to their health care needs. They also did not allow for any referrals of the patients. They formulated the problem as a zero-one integer programming problem. They did not propose any special method, that exploits the structure of the problem, to solve it.

In this paper, to keep the exposition simple, we consider a health care delivery system location–allocation problem that consists of p_1 health centers and p_2 hospitals among n $(\geq p_1 + p_2)$ potential locations when the objective is to minimize the total weighted travel distance. (The extension to three types of facilities will appear later.) We allow the referral of θ $(0 \leq \theta \leq 1)$ proportion of patients from a health center to a hospital. We shall refer to this problem as a 2-hierarchal location–allocation problem. It may be noted that if the health care delivery system consisted only of health centers or only of hospitals, the problem would reduce to the well known p-median problem. Recently, Narula, Ogbu and Samuelsson [10] proposed an efficient algorithm to solve the p-median problem.

In section 2, we extend the notion of a median of a graph to a hierarchal median set of a graph and show that the 2-hierarchal location–allocation problem can be analyzed as a hierarchal median set problem. We give a mathematical programming formulation of the problem in section 3. In section 4, we discuss various heuristic procedures followed by some computational experience and conclusions in sections 5 and 6, respectively.

2. HIERARCHAL MEDIAN SET

Consider a connected graph G that consists of n vertices indexed by $i = 1, \ldots, n$. A weight W_i is attached to a vertex i. Let $d(r, s)$ be the shortest distance between any two points r and s in the graph. We associate a vertex with a population center and the length of the arc between two vertices with the shortest distance between the corresponding population centers. Let θ proportion of patients be referred from a health center to a hospital.

Definition 1

A point $\{r^0, s^0\}$ is a 2-hierarchal absolute median set of G if for any other set $\{r, s\}$

$$\sum_{i=1}^{n} W_i \min \{d(i, r^0) + \theta d(r^0, s^0), d(i, s^0)\} \leq \sum_{i=1}^{n} W_i \min \{d(i, r) + \theta d(r, s), d(i, s)\}.$$

If $\{r^0, s^0\}$ represents a vertex set, then a 2-hierarchal absolute median set is called a 2-hierarchal vertex median set.

Theorem 1

If a 2-hierarchal absolute median set of a graph G is restricted to the vertices and the arcs connecting the vertices, then at least one 2-hierarchal vertex median set is a 2-hierarchal absolute median set.

The proof of the theorem is given in Ogbu [12] and follows from the proof of Hakimi [6] for the absolute median of a graph.

Thus a 2-hierarchal location–allocation problem of a health care delivery system that consists of one health center and one hospital is equivalent to a 2-hierarchal vertex median problem.

Definition 2

An ordered pair of sets $\{R^0_{p_1}, S^0_{p_2}\}$ is a 2-hierarchal (p_1, p_2) absolute median set of G if for any other ordered pair set $\{R_{p_1}, S_{p_2}\}$

$$\sum_{i=1}^{n} W_i \min \left\{ \min_{r^0 \in R^0_{p_1}, s^0 \in S^0_{p_2}} (d(i, r^0) + \theta d(r^0, s^0)), \min_{s^0 \in S^0_{p_2}} d(i, s^0) \right\} \leq \sum_{i=1}^{n} W_i \min \left\{ \min_{r \in R_{p_1}, s \in S_{p_2}} (d(i, r) + \theta d(r, s)), \min_{s \in S_{p_2}} d(i, s) \right\}$$

where R_{p_1} and S_{p_2} denote sets of cardinality p_1 and p_2, respectively.

If $\{R^0_{p_1}, S^0_{p_2}\}$ represent an ordered pair of vertex set, then a 2-hierarchal (p_1, p_2) absolute median set of G is called a 2-hierarchal (p_1, p_2) vertex median set.

Theorem 2

If a 2-hierarchal (p_1, p_2) absolute median set is restricted to the vertices and arcs of G, then at least one 2-hierarchal (p_1, p_2) vertex median set is a 2-hierarchal (p_1, p_2) absolute median set.

The proof of the theorem is given in Ogbu [12]. A 2-hierarchal location–allocation problem of a health care delivery system that consists of p_1 health centers and p_2 hospitals is equivalent to a 2-hierarchal (p_1, p_2) vertex

Vol. 7
1979

median problem. By theorem 2, a 2-hierarchal (p_1, p_2) median set can be obtained by evaluating $n!/[p_1!\ p_2!(n - p_1 - p_2)!]$ possible vertex combinations. This can be a very large combinatorial problem. The enormity of the problem can be appreciated by comparing the number of vertex combinations of a 2-hierarchal (p_1, p_2) median problem to an equivalent p-median problem.

$$\text{This ratio is } \frac{(p_1 + p_2)!}{p_1!p_2!}.$$

3. PROBLEM FORMULATION

Consider the location–allocation problem of p_1 health centers and p_2 hospitals among n $(\geq p_1 + p_2)$ population centers such that the total weighted travel distance is minimized when θ proportion of the patients at a health center are referred to a hospital.

For given facilities locations, the optimal allocation of patients to the health facilities, when the facilities are uncapacitated, is a simple weighted path problem. The optimal allocation of patients to the health facilities is as follows:

1. For each population center i with no health facility allocate the patients to a health facility such that the travel distance is minimum of the travel distance to the nearest hospital and the expected travel distance to a hospital via a health center.

2. For each location i with a health center allocate the referred patients to the nearest hospital.

3. For each location i with a hospital, allocate the patients to the hospital.

From the preceding (optimal) allocations, it is clear that it may be optimal to allocate patients to a health facility other than the nearest facility even when the facilities are uncapacitated.

In the mathematical programming formulation of the capacitated 2-hierarchal (p_1, p_2) median problem we shall use the following notation:

W_i = number of patients at location i,
X_{ij}^{01} = number of patients at location i, with no facility, allocated to location j with a health center,
X_{ij}^{02} = number of patients at location i, with no facility, allocated to location j with a hospital,
X_{ij}^{12} = number of patients at location i, with a health center, allocated (referred) to location j with a hospital,
$Y_j^1 = \begin{cases} 1, & \text{if a health center is located at } j, \\ 0, & \text{otherwise,} \end{cases}$
$Y_j^2 = \begin{cases} 1, & \text{if a hospital is located at } j, \\ 0, & \text{otherwise,} \end{cases}$
p_1 = number of health centers,
p_2 = number of hospitals,
θ = fraction of patients at a health center referred to a hospital, $0 \leq \theta \leq 1$,
C^1 = capacity of a health center,
C^2 = capacity of a hospital,
d_{ij} = minimum travel distance between locations i and j.

The objective is to minimize the total weighted travel distance.

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}(X_{ij}^{01} + X_{ij}^{02} + X_{ij}^{12}).$$

Subject to:

The patients at each population center are allocated to a health facility.

$$\sum_{j=1}^{n} (X_{ij}^{01} + X_{ij}^{02}) = W_i, \quad i = 1, \ldots, n. \tag{1}$$

A fraction θ of the patients allocated to a health center at i are referred to a hospital.

$$\sum_{\substack{j=1 \\ i \neq j}}^{n} X_{ij}^{12} = \theta \sum_{j=1}^{n} X_{ji}^{01}, \quad i = 1, \ldots, n. \tag{2}$$

Health center allocations are made only to locations with health centers

$$\sum_{i=1}^{n} X_{ij}^{01} \leq C^1 Y_j^1, \quad j = 1, \ldots, n. \tag{3}$$

Hospital allocations are made only to locations with hospitals.

$$\sum_{i=1}^{n} X_{ij}^{02} + \sum_{\substack{i=1 \\ i \neq j}}^{n} X_{ij}^{12} \leq C^2 Y_j^2, \quad j = 1, \ldots, n. \tag{4}$$

Exactly p_1 health center locations are selected.

$$\sum_{j=1}^{n} Y_j^1 = p_1. \tag{5}$$

Exactly p_2 hospital locations are selected.

$$\sum_{j=1}^{n} Y_j^2 = p_2. \tag{6}$$

Each location can have at most one facility

$$Y_j^1 + Y_j^2 \leq 1, \quad j = 1, \ldots, n. \tag{7}$$

$$0 \leq X_{ij}^{01} \leq W_i; \quad 0 \leq X_{ij}^{02} \leq W_i; \quad 0 \leq X_{ij}^{12} \leq \theta C^1,$$

$$Y_j^1 = 0 \text{ or } 1; \quad Y_j^2 = 0 \text{ or } 1.$$

This is a mixed integer programming (MIP) problem with $2n$ zero–one integer variables, $3n^2$ continuous variables and $5n + 2$ constraints. To solve the problem as an uncapacitated problem, we simply make C^1 and C^2 suitably large.

4. HEURISTIC METHODS

Although the problem can be solved by any of the available MIP codes, this is not advisable as the problem size is large even for a moderate value of n. Since we can restrict our search for an optimum solution to the population centers only (theorem 2 section 2), we propose the following five heuristic procedures:

Forward p-*median heuristic*

First we solve a $(p_1 + p_2)$-median problem. These $p_1 + p_2$ medians (locations) are used to locate the health centers and hospitals. The rationale is that we expect a median (location) in the 2-hierarchal median set to be a median in the $(p_1 + p_2)$-median problem. Next, we use the $p_1 + p_2$ median locations to solve a p_2–median problem. These p_2 medians are used to locate hospitals and the remaining p_1 medians (of the $p_1 + p_2$ medians) to locate health centers. The rationale is that the p_2 medians are the most important with regard to minimizing the total weighted travel distance, and hence should be used to locate hospitals. This procedure was first proposed by Harvey, Hung and Brown [7] for $\theta = 0$. We modify the procedure to introduce θ as follows:

Let N denote the set of population centers, i.e. $N = \{i | i = 1, 2, \ldots, n\}$. First solve a $(p_1 + p_2)$-median problem. Let $V_{(p_1 + p_2)}$ denote the $p_1 + p_2$ medians and let

$$F_j(i) = \left\{ i \in N \,|\, d_{ij} = \min_{j \in V_{(p_1 + p_2)}} d_{ij} \right\},$$

i.e. $F_j(i)$ is the set of population centers for which j is the absolute median. For a location–allocation problem, $F_j(i)$ is the set of population centers served by a facility at j. For every $j \in V_{(p_1 + p_2)}$, compute

$$W_j^i = W_j + \sum_{i \in F_j(i)} \theta W_i, \quad 0 \leq \theta \leq 1. \tag{8}$$

Solve a p_2-median problem in the $p_1 + p_2$ median locations using the weights in (8). The p_2 hospitals are located at these p_2 medians and p_1 health centers are located at the remaining p_1 median locations.

Backward p-*median heuristic*

This procedure takes a reverse approach to that of the forward procedure. First, we solve a p_2-median problem. These p_2 medians are used to locate hospitals. Next, we solve a p_1-median problem for the points not in the p_2 medians. These p_1 medians are used to locate health centers.

Add heuristic

The add heuristic described here is essentially an extension of the add heuristic of Kuehn and Hamburger [8] for $k = 1$, i.e. the p-median problem. The procedure essentially is to select one facility location at a time until all the facility locations have been selected. For a 2-hierarchal (p_1, p_2) median problem, one can devise a number of add heuristics. However, we used the following procedure: select one location at a time for a hospital using other locations as health centre locations, until p_2 locations have been selected. Then select one location at a time for a health center, until p_1 locations have been selected.

Drop heuristic

This heuristic is essentially an extension of the drop heuristic of Feldman, Lehrer and Ray [3] for the p-median problem and takes a reverse approach to that of the add heuristic. The procedure is similar to the drop heuristic proposed by Scott [14] for the factory, warehouse and retail-outlet location–allocation problem.

The procedure starts with the assumption that a health center and a hospital exist at each location. One hospital is dropped at a time using other locations for health centers until p_2 hospitals are left. Then one health center is dropped at a time until p_1 health centers are left.

Vol. 7
1979

Greedy interchange heuristic

This heuristic starts with any p_1 locations for health centers and p_2 locations for hospitals and then seeks to improve upon the solution. An improvement may be achieved by an interchange between a facility location and a non-facility location and/or a switch between a health center and a hospital location. This procedure is essentially an extension of the well known heuristic proposed by Tietz and Bart [15] for the p-median problem. However, an important difference between the two procedures is the criterion used for the interchange.

Let Z_{ij}^{st}, ($s = 0$, 1 and $t = 1$, 2) be the improvement, if any, obtained by substituting (intrachanging/interchanging) a type s facility at location i for a type t facility at location j ($s = 0$ means no facility). For each i, Tietz and Bart retain an interchange for max Z_{ij}^{st}. This criterion is similar to the variable selection criterion used in the simplex algorithm. Our criterion is to retain an interchange whenever Z_{ij}^{st} is strictly positive, i.e. whenever an improvement can be obtained. Because of this criterion, we call our procedure greedy procedure. This criterion is similar to a chronological selection of variables in the simplex method. The motivation for the greedy heuristic was primarily computational.

Observe that these heuristic procedures provide an upperbound Z_U on the solution.

6. COMPUTATIONAL EXPERIENCE

To gain computational experience with the proposed procedures a few generated problems were solved by the Forward p-median Heuristic, the Backward p-median Heuristic, the Add Heuristic, the Drop Heuristic, and Greedy Interchange Heuristic. The Forward and the Backward p-median Heuristics use the p-median algorithm given in Narula, Ogbu, and Samuelsson [10]. Five problems of each size (n, p_1, p_2, θ) were generated. The W_i were uniformly distributed between 1 and 20, and the coordinates of the demand locations were uniformly distributed between 1 and 200.

To compare the quality of upper bounds Z_U provided by the proposed heuristic procedures, we need a standard of comparison. However, in the absence of an obvious standard—optimal solutions to the problems, we compared the upper bounds to a lower bound Z_L. The lower bound was obtained by solving the lagrangean relaxation of the problem by a subgradient optimization procedure. (The details of this procedure are given in Narula and Ogbu [9].) To compare the heuristic procedures in terms of the quality of the solution, for each heuristic we computed $R = (1 - Z_L/Z_U)\,100$, the percentage difference between the upper and the lower bounds. Table 1 gives the average values of R for various heuristics and problem sizes. Thus in terms of the quality of upper bounds, the heuristics can be arranged

TABLE 1. AVERAGE PERCENT DIFFERENCE BETWEEN AN UPPER BOUND AND A LOWER BOUND

n	p_1	p_2	θ	Greedy interchange heuristic	Forward p-median heuristic	Backward p-median heuristic	Add heuristic	Drop heuristic
10	1	2	0.4	18	22	29	24	26
			0.2	8	14	22	15	16
			0.1	7	7	28	22	16
10	3	2	0.4	24	24	27	27	31
			0.2	24	28	32	29	34
			0.1	14	16	29	17	28
20	2	4	0.4	16	20	24	20	24
			0.2	11	11	22	13	19
			0.1	7	11	23	20	17
20	6	4	0.4	30	32	35	35	34
			0.2	24	26	30	30	35
			0.1	13	21	27	24	27
20	2	8	0.4	8	8	18	15	47
			0.2	4	6	17	13	33
			0.1	4	6	14	13	36
20	6	8	0.4	20	22	26	26	52
			0.2	9	9	18	15	43
			0.1	7	7	18	14	42

TABLE 2. AVERAGE COMPUTATIONAL TIME* (CPU SECONDS)

n	p_1	p_2	θ	Greedy interchange heuristic	Forward p-median heuristic	Backward p-median heuristic	Add heuristic	Drop heuristic
10	1	2	0.4	0.21	0.21	0.20	0.09	0.46
			0.2	0.10	0.19	0.09	0.09	0.45
			0.1	0.11	0.21	0.24	0.09	0.46
10	3	2	0.4	0.19	0.17	0.29	0.11	0.45
			0.2	0.20	0.08	0.31	0.11	0.45
			0.1	0.20	0.11	0.33	0.11	0.45
20	2	4	0.4	1.65	1.41	1.52	1.35	6.89
			0.2	1.88	0.47	1.60	1.33	6.99
			0.1	1.53	0.72	1.66	1.35	6.93
20	6	4	0.4	2.85	0.70	1.77	1.53	6.75
			0.2	3.19	0.84	1.57	1.51	6.63
			0.1	3.26	1.31	1.64	1.53	6.69
20	2	8	0.4	2.27	0.42	1.02	2.09	5.25
			0.2	2.36	0.57	0.88	2.07	5.11
			0.1	2.25	0.55	0.79	2.07	5.13
20	6	8	0.4	3.56	0.34	1.41	2.16	5.08
			0.2	3.13	0.53	1.36	2.17	5.22
			0.1	2.84	0.54	0.79	2.17	5.13

* Cyber 173 Computer.

Vol. 7
1979

in order of decreasing perference as: (1) Greedy Interchange Heuristic, (2) Forward p-median Heuristic, (3) Add Heuristic, (4) Backward p-median Heuristic, and (5) Drop Heuristic. As expected, the percent difference R decreases as θ decreases.

The average computational times required to solve a problem by the five heuristic procedures are given in Table 2. In terms of the computational time required, the methods can be arranged in a decreasing order of preference as: (1) Forward p-median Heuristic, (2) Backward p-median Heuristic, (3) Add Heuristic, (4) Greedy Interchange Heuristic, and (5) Drop Heuristic.

7. CONCLUSIONS AND REMARKS

The hierarchal nature of facility systems has been only marginally considered in published research on location problems. We have highlighted the importance and relevance of this feature in such problems.

For a class of hierarchal systems, we have discussed and formulated the problem in relation to a health care delivery system that consists of health centers and hospitals. From among the five heuristics proposed, we recommend the Greedy Interchange Heuristic and the Forward p-median Heuristic procedures in terms of their quality of upper bounds and the computational times required to solve a problem.

Although we have discussed the problem in relation to a health care delivery system, the results of this paper are applicable to a number of other systems with similar features. Some examples of other such systems are: higher education (4-year colleges, universities), banks (drive-ins, branch offices), post offices (drug store outlets, regional branches).

ACKNOWLEDGEMENTS

We would like to thank Dr S Kaufman and the late Dr HM Samuelsson for helpful discussions. We also thank the Department of Industrial Engineering of SUNY at Buffalo for the computer time.

REFERENCES

1. BRYANT J (1969) *Health and the Developing World.* Cornell University Press, Ithaca, USA.
2. CALVO AB & MARKS DH (1973) Location of health care facilities: an analytic approach. *Socio-Econ. Plann. Sci.* **7**, 407–484.
3. FELDMAN E, LEHRER FA & RAY TL (1966) Warehouse locations under continuous economies of scale. *Mgmt Sci.* **12**, 670–684.
4. FENDALL NRE (1963) *Planning health services in developing countries.* Public Health Reports, **78**, 977–988.
5. GISH O (1971) *Health, Manpower and the Medical Auxiliary.* Intermed. Technol. Dev. Group, London, Engl.
6. HAKIMI SL (1964) Optimum locations of switching centers and the absolute centers and medians of a graph. *Ops Res.* **12**, 450–459.

7. HARVEY MS, HUNG MS & BROWN JR (1973) The application of a *p*-median algorithm to the identification of nodal hierarchies and growth centres. *Econ. Geogr.* **50**, 187–202.
8. KUEHN AA & HAMBURGER MJ (1963) A heuristic program for locating warehouses. *Mgmt Sci.* **9**, 643–666.
9. NARULA SC & OGBU UI (1978) *A 2-hierarchal location–allocation problem.* Res. Rep. No. 37-78-P2, School of Management., Rensselaer Polytechnic Institute Troy, New York.
10. NARULA SC, OGBU UI & SAMUELSSON HM (1977) An algorithm for the *p*-median problem. *Ops Res.* **25**, 709–713.
11. NAVARRO V (1972) Health services in Cuba. *New Engl. Jl Med.* **19**, 954–959.
12. OGBU UI (1978) *Uncapacitated k-hierarchal system location problems.* Ph.D. Diss. State University of New York at Buffalo, Buffalo, New York.
13. SCHULTZ GP (1970) The logic of health care facility planning. *Socio-Econ. Plann. Sci.* **4**, 383–393.
14. SCOTT AJ (1971) Operational analysis of nodal hierarchies in network systems. *Ops. Res. Q.* **22**, 25–36.
15. TIETZ MB & BART P (1968) Heuristic method for estimating the generalized vertex median of a weighted graph. *Ops Res.* **16**, 955–961.

ADDRESS FOR CORRESPONDENCE: *Professor Subhash C Narula, School of Management, Rennsselaer Polytechnic Institute, Troy, New York* 12181, *USA.*

ol. 7
979

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 2. pp. 145 to 151
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0401-0145$02.00/0

Solving the Two-Stage Cutting Stock Problem

ROBERT W HAESSLER

The University of Michigan, USA

(*Received July* 1978; *in revised form October* 1978)

This paper describes an important class of cutting stock problems not previously discussed in the literature. The problem is one of determining the patterns to be used in a two-stage cutting process with restrictions imposed on the locations of cuts in the first stage. A particularly difficult version of this problem from the plastic film industry is presented and solved.

INTRODUCTION

OVER THE past 20 years, a substantial amount of work has been done on one and two-dimensional cutting stock problems. A number of early authors [1, 6–8] recognized that the roll trim problem in the paper industry could be formulated as a linear programming model. The pioneering computational work by Gilmore and Gomory [2, 3] on delayed pattern generation techniques made it practical to find linear programming solutions to large one-dimensional problems. Follow-up research by Pierce [9] focused on the use of this algorithm in the paper industry to solve the roll trim problem. Gilmore and Gomory [4] subsequently extended their work to higher dimensional problems such as the cutting of rectangles from rectangles, a common problem in the glass industry. They overcame the difficulty of generating patterns in higher dimensions by limiting themselves to patterns with guillotine cuts (once a cut is begun it must be continued through to the other side of the material in a straight line). As a result the problem of generating a higher dimensional cutting pattern can be handled by sequentially solving a number of one-dimensional problems. Their multi-stage approach in the two-dimensional case focuses first on cutting patterns for strips of known width and length and then on ways these strips can be combined to form two-dimensional patterns.

In a number of industries, such as paper and film, there is an important class of problems that falls in between the well known one and two-dimensional problems. These are referred to here as two-stage cutting stock problems because the material is processed through two distinct cutting operations. It differs from the multi-stage cutting problems discussed by Gilmore and Gomory because the first and second stage cuts are parallel to one another. This problem has not been discussed in the literature even though it occurs in a number of industries. This two-stage cutting may be necessary because of the nature of the production process and the type of equipment available or to give flexibility so that orders that vary in some characteristic other than width can be combined across the width of the primary production process. Regardless of the reason, production rolls are first slit into master roll widths at the first stage and the master roll widths are then slit to the ordered sizes at the second stage. Limitations in the master roll sizes that can be processed at the second stage make it necessary to impose additional restrictions on the pattern generating process relative to the well known one-dimensional roll trim problem.

GENERAL PROBLEM CONSIDERATIONS

Consider the production process sketched in

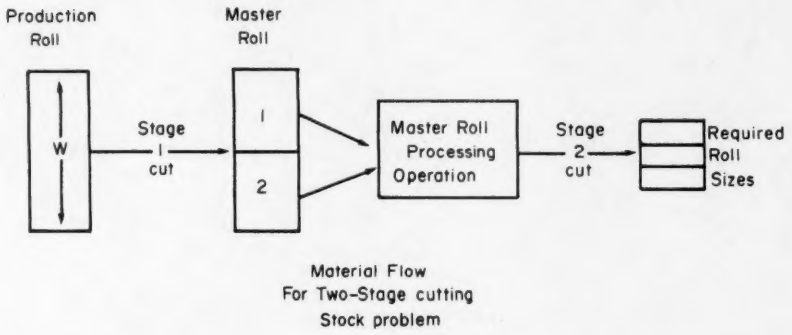


FIG. 1. *Material flow for two-stage cutting stock problem.*

Fig. 1. Let W be the width of production rolls available to fill a known set of order requirements for rolls of some product. In the first cutting operation the production rolls are cut to obtain master rolls. After some additional processing these master rolls are cut to obtain the required roll sizes. These downstream operations place limits on the widths of the master rolls to be cut at the first stage. The problem is to specify how to cut the production rolls and resulting master rolls to meet the order requirements at minimum cost. If the production rolls could be cut in a single operation to obtain the required roll sizes, this would be a one-dimensional trim loss minimization problem. It is the size limitations imposed by the master roll processing operation as well as the costs incurred at this operation that make this problem unique.

Any cutting stock problem can be formulated as a trim loss minimization problem as follows:

$$\min \quad \Sigma_p x_p \tag{1}$$

$$\Sigma_p A_{ip} x_p = R_i \quad \text{all} \quad i \tag{2}$$

$$x_p \geq 0 \tag{3}$$

where

x_p is the usage of pattern p.

A_{ip} is the number of times size i appears in pattern p,

and

R_i is a measure of the required quantity for size i.

The differences in cutting stock problems can be seen only by defining what constitutes a feasible cutting pattern. In the one-dimensional problem with order sizes W_i for $i = 1, \ldots, m$, solved by Gilmore and Gomory, a cutting pattern is a set of values A_{ip} for $i = 1, . , . , m$ that satisfy the following conditions

$$\Sigma_i A_{ip} W_i \leq W \tag{4}$$

$$A_{ip} \geq 0, \quad \text{integer.} \tag{5}$$

In the two-stage cutting problem considered here the A_{ip} must satisfy (4) and (5) but also must be able to be partitioned into elements A_{ipk} which satisfy the following conditions

$$A_{ip} = \Sigma_k A_{ipk} \quad \text{for all} \quad i, p \tag{6}$$

$$\Sigma_i A_{ipk} W_i \leq W M_k \quad \text{for all} \quad k, \tag{7}$$

$$A_{ipk} \geq 0, \quad \text{integer} \tag{8}$$

where

A_{ipk} is the number of times sizes i appears in master roll k for pattern p

and

WM_k is the maximum width of master roll k.

In short, each pattern for the full width of the production roll can be thought of as defining a cutting stock problem with the master roll sizes defining the stock sizes available. This embedded cutting stock problem must have a feasible solution in order for the cutting pattern to be admissible. For example suppose that W, the maximum production width, is 200 inches and that each production roll must be slit into two master rolls neither of which can exceed 110 inches. A pattern containing eight 25-inch rolls is satisfactory because two master rolls each containing four 25-inch rolls can be cut. However, a pattern containing seven 28-inch rolls would not be acceptable if only two master rolls are to be cut because four times 28 is more than 110. If more than two master rolls can be cut, this pattern could be

Vol. 7
1979

used. However, the cost of processing three master rolls may make this pattern less desirable than one with the same trim loss that requires only two master rolls.

In addition to the above pattern restrictions, there may be additional differences relative to the one-dimensional roll trim problem. In some cases the first cutting operation may be done in such a way that the master rolls that are made side by side need not be the same diameter. This permits the possibility of combining orders of different diameters in the same pattern. In other cases, the master rolls may eventually be processed into sheets and therefore there may be some flexibility in determining the diameter of the master rolls and the diameters used may differ from pattern to pattern. Under these circumstances the second dimension of the order (i.e. roll diameter) does come into play and takes the problem beyond the one-dimensional problem. However, it is not really a two-dimensional problem because the length dimension of the stock size is not fixed. The handling of these considerations will differ from application to application. This will be demonstrated more fully in the film slitting problem described below.

FILM SLITTING PROBLEM

A particularly interesting example of a two-stage cutting stock problem is found in the plastic film industry. In this case the film is slit at the end of the extruder on which it is formed. By slitting on the extruder the master rolls produced are independent of one another along the length of production, and therefore can be produced at different diameters. The master rolls are then taken immediately to the second stage slitters where they are slit to the ordered widths. The width capacity of the second stage slitters limits the master roll widths that can be produced.

Because the master roll diameters can be separately controlled, it is possible to consider problems that have mixed diameters or lineal footage requirements. This cannot be done in one-dimensional trim problems. The only requirement is that the orders to be slit from any master roll must all be of the same diameter. This simply puts an additional restriction on the conditions a pattern must satisfy to be feasible. The price that must be paid to obtain this flexibility with master roll diameters is that the process must be shut down to change the slit point setting. This introduces a fixed charge that is incurred each time the master roll widths must be changed.

Note that it is possible to generate scrap along the length of production as well as across the width. Suppose that 100,000 feet of film is to be produced for one slit point setting on the extruder. On one side of the slit point it is possible to produce five master rolls each of which has a length of 20,000 feet. Suppose that on the other side of the slit point the master rolls required must have 16,000 feet of material. Only six of these can be produced and there will be 4,000 feet of material left over that must be scrapped or held in inventory until some use can be found for it.

To formulate this problem let

W_i	be the roll width order i
L_i	be the lineal footage for each roll for order i
RL_i	be the lower bound on the rolls required to satisfy order i
RU_i	be the upper bound on the rolls required to satisfy order i
W	be the width of the extruder
WM_k	be the maximum width for master roll k
C_1	be the value of a unit area of trim loss
C_2	be the cost of shutting down the extruder to change the slit point setting

The problem is to produce the order requirements within the tolerances specified, at a minimum combined cost of trim loss and downtime. The problem can be formulated mathematically as follows:

$$\min C_1(W\Sigma_j X_j - \Sigma_i Y_i L_i W_i) + C_2 \Sigma_j \delta(X_j) \tag{9}$$

subject to

$$RL_i \leq Y_i \leq RU_i \quad \text{for all } i, \tag{10}$$

$$Y_i = \Sigma_j \Sigma_k \Sigma_p A_{ijkp}[X_{jkp}/L_i] \quad \text{for all } i, \tag{11}$$

$$X_j \geq \Sigma_p X_{jkp} \quad \text{for all } j \text{ and } k, \tag{12}$$

and

$$X_j, X_{jkp} \geq 0 \tag{13}$$

where the remaining symbols are defined as follows:

Y_i—number of rolls produced for order i.

X_j—number of feet of production for the jth set of master roll widths. (A set of master roll widths is determined by the slit point setting at the extruder.

$\delta(X_j) = 1$ if $X_j > 0$, 0 otherwise

A_{ijkp}—number of times order i appears in pattern p for master roll k and slit point setting j.

X_{jkp}—number of feet of production of master roll k processed according to pattern p for slit point setting j.

$[X_{jkp}/L_i]$—largest integer less than or equal to X_{jkp}/L_i.

To aid in the discussion, let Z_{jk} be the width of master roll k for the jth set of master roll sizes. It follows that:

$$\Sigma_k Z_{jk} = W \quad \text{for all} \quad j, \tag{14}$$

$$Z_{jk} \leq WM_k \quad \text{for all} \quad j \text{ and } k, \tag{15}$$

$$\Sigma_i A_{ijkp} W_i \leq Z_{jk} \quad \text{for all} \quad j, k \text{ and } p, \tag{16}$$

and

$$A_{ijkp} \geq 0, \quad \text{integer} \tag{17}$$

This formulation contains some extreme combinatorial aspects and is far beyond the capability of currently available optimization procedures. Therefore attention is directed toward the problem of finding 'good' feasible solutions. The primary objective is to develop a procedure that generates better solutions than are obtainable by currently available manual methods. Although the mathematical model can not be solved for an optimal solution, it does provide a basis for testing solution feasibility and evaluating solution quality regardless of how the solution is generated.

SOLUTION PROCEDURE

A computer program that employs a number of heuristics in conjunction with a linear programming algorithm was developed to solve this problem for a film producer [5]. A detailed description of the solution procedure follows:

1. Compute the total lineal footage for each order.

$$F_i = R_i L_i \tag{18}$$

This then becomes the required quantity in a linear programming problem.

$$\text{minimize} \quad \Sigma_p X_p \tag{19}$$

$$\text{subject to} \quad \Sigma_p A_{ip} X_p = F_i \quad \text{for} \quad i = 1 \ldots m, \tag{20}$$

$$X_p \geq 0 \tag{21}$$

where

X_p—number of feet of production to process using pattern p.

A_{ip}—number of times size i appears in pattern p.

2. Solve the problem given above using a Gilmore–Gomory [1, 2] type linear programming algorithm. This algorithm is based on the principle of delayed pattern generation. All possible patterns are not generated at the outset. Patterns that will improve the solution are generated as needed by solving an associated problem to find that pattern that will give the greatest incremental improvement. This pattern generation process is restricted such that only patterns that can be partitioned to meet the slitter width and length restrictions are considered. This insures solution feasibility across the width of the extruder. A feasible cutting pattern in a classical roll slitting problem can be defined by elements A_{ip} such that:

$$\Sigma_i A_{ip} W_i \leq W \tag{22}$$

$$A_{ip} \geq 0, \text{ integer} \tag{23}$$

In this case, the A_{ip} must also be partitioned into elements A_{ikp}, the number of times size i appears in master roll k for pattern p. The following restrictions must be met:

$$\Sigma_k A_{ikp} = A_{ip} \quad \text{all} \quad i \text{ and } p \tag{24}$$

$$\Sigma_i A_{ikp} W_i \leq WM_k \quad \text{all} \quad k \tag{25}$$

In addition, because of slitter limitations, all the orders contained in a master roll pattern must have the same lineal footage.

The solution to the linear programming problem minimizes the trim loss and the requirements in terms of footage, F_i, for each order are met exactly. However, this solution is generally of no use because it usually will require a substantial amount of splicing to obtain rolls of the desired lineal footage. In most situations, splicing is not even allowed, therefore the pattern production quantities must be determined for each master roll pattern as multiples of the lineal footage of the

Vol. 7
1979

involves finding the slit point settings and grouping the patterns that can be produced at the same setting.

3. Determine the number and location of the slit point settings for the patterns found in 2. This can be done by computing the minimum and maximum possible width for the widest master roll in each pattern, finding the smallest of the maximums, and taking that as a slit point setting and assigning as many full width patterns as possible to that setting. The minimum and maximum widths will not be the same if $W > \Sigma_i A_{ip} W_i$, there is trim loss in the pattern. This procedure is repeated until all patterns are assigned to a setting. At this point X_p becomes X_{jp}, where j identifies the slit point setting. This procedure will give the minimum number of settings when there is only one slit point at the extruder. When there is more than one slit point, the procedure must be applied to each slit point. It is no longer possible to guarantee that the minimum number will be found.

4. For each master roll width in each pattern, find a length that will yield an integer number of rolls. This is done by dividing X_{jp} by L_i, the length of the orders in that master roll and rounding to the nearest integer. Now X_{jp} becomes X_{jkp} where X_{jkp} is an integer multiple of the order length that is being made in that master roll pattern.

5. Compute the total length for each master roll setting.

$$X_j = \max_k (\Sigma_p X_{jkp}) \tag{26}$$

and compute the number of each size roll that is being produced.

At this point, there are three potential difficulties with the solution.

(a) The amount produced for an order may be less than the lower bound of the requirement as a result of step 4.

(b) There may be excessive trim loss along the length for a given master roll width. The unused portion of master roll k for slit point setting j is $X_j - \Sigma_p X_{jkp}$.

(c) The solution may involve too many slit point settings.

These three potential problems are handled as follows:

6. Increase the production length of the first master roll pattern found that contains a size for which the production quantity is insufficient to meet the lower bound, provided that it does not cause some other item to be produced at a level above its upper bound. This process is repeated until either all the items fall within the production limits or until no further progress can be made. Each time the length of a master roll pattern is increased, the values of X_j and X_{jkp} must be changed accordingly.

7. Check each slit point setting to determine if there is a significant length of unused material for any master roll. If there is a loss for master roll t, do the following in the order given.

(a) Reduce the usage X_{jkp} of patterns for master rolls $k \neq t$ to reduce the $\Sigma_p X_{jkp}$ and thereby be able to reduce X_j. This reduction is done only if the order requirements are not violated.

(b) Increase the usage X_{jkp} of some pattern for master roll $k = t$ to make use of the unutilized material. Once again this is done only if the upper bound for each order is not violated.

At this point a feasible solution should have been found, and the only question remaining is whether it is economically desirable to reduce the number of slit point settings. Reducing the number of slit point settings will generally increase the trim loss (it cannot reduce the trim loss except for differences due to rounding). The issue is whether or not the increase in trim loss is less than the gain in production obtained by eliminating a slit point change.

8. Determine the slit point setting which has the lowest usage X_j. Eliminate it from the set of slit point settings than can be used during the next solution to the problem. This can be done by specifying a set of maximum master roll widths Z_{jk} such that:

$$\Sigma_k Z_{jk} = W \quad \text{for all} \quad j \tag{27}$$

The entire problem is then resolved. The only difference is that step 3 can be eliminated. The patterns are assigned to one of the allowed slit point settings at the time they are generated. This process repeats until the increased cost of trim loss due to reducing the number of

orders scheduled. The first step in this process master roll settings more than offsets the decrease in the cost of downtime made possible by reducing the number of master roll settings.

The tradeoff between trim loss and downtime is handled by specifying the number of feet of production lost while changing the master roll settings. This is multiplied by W to find the area of production lost. This is then added to the actual trim loss to find the effective loss of a solution.

EXAMPLE

An example problem is given in Table 1. The extruder in this example is capable of producing film with a maximum usable width of 124 inches. The two second stage slitters that process the master rolls have maximum widths of 76 and 55 inches. This means that as the film comes off the extruder it must be slit exactly once at a point that is within 76 inches of one side and 55 inches of the other so the resulting master rolls can be processed through the two second stage slitters. It is not possible to produce two 62-inch width master rolls because they would both have to go the larger slitter and the slitter processing speed is not sufficient to keep up with the extruder under these conditions.

The solution to this problem is presented in Table 2. The lost production of a slit point change was specified as 30,000 feet. The result-

TABLE 1. SAMPLE PROBLEM

Production data

Extruder width—124 inches
Number of slit points—1
Width of slitter 1—76 inches
Width of slitter 2—55 inches

Roll order requirements

Quantity	Width	Lineal footage/roll
7	55.0000	20000.
3	39.0000	20000.
4	36.5000	35000.
1	36.0313	10000.
36	26.0000	10000.
270	24.5000	15000.
4	21.0000	15400.
4	21.0000	15400.
3	19.0000	20000.
7	18.2500	20000.
4	17.8750	20000.

TABLE 2. PROBLEM SOLUTION

Master roll widths
1 75 inches
2 49 inches
Total Production—1,141,600 feet
Yield—94.5%

Master roll 1 length 20000.
1 55.0000 20000.
1 17.8750 20000.
Master roll 2 length 60000.
2 24.5000 15000.

Pattern 2

Master roll 1 length 60000.
1 55.0000 20000.
1 18.2500 20000.
Master roll 2 length 60000.
1 39.0000 20000.

Pattern 3

Master roll 1 length 70000.
2 36.50000 35000.
Master roll 2 length 60000.
2 24.50000 15000.

Pattern 4

Master roll 1 length 20000.
1 55.0000 20000.
1 18.2500 20000.
Master roll 2 Length 15000.
2 24.5000 15000.

Pattern 5

Master roll 1 length 120000.
2 26.0000 10000.
Master roll 2 length 90000.
2 24.5000 15000

Pattern 6

Master roll 1 length 720000.
3 24.5000 15000.
Master roll 2 length 735000.
2 24.5000 15000.

Pattern 7

Master roll 1 length 60000.
2 26.0000 10000.
Master roll 2 length 61600.
2 21.0000 15400.
0 21.0000 15400.

Pattern 8

Master roll 1 length 60000.
1 55.0000 20000.
1 19.0000 20000.
Master roll 2 length 60000.
1 18.2500 20000.
1 17.8750 20000.

Pattern 9

Master roll 1 length 10000.
1 36.0313 10000.
1 26.0000 10000.
Master roll 2
1600 feet of scrap

Vol. 7
1979

ing solution has only one slit point setting and an overall yield of 94.5% on scheduled production of 1,141,600 feet. There are 9 master roll patterns for the 76 inch slitter and 8 for the 55 inch slitter. There are 1,600 feet of unutilized material on the master rolls going to the first slitter. This is treated as scrap. The master roll patterns are given in pairs to indicate the manner in which they were generated. The differences in usage for corresponding master roll patterns indicates the amount of adjusting that is done to meet the order requirements and avoid leaving unutilized material. The only deviations from the ordered quantities are the following overruns, all within the 10% deviation allowed.

1 55″

1 26″

2 24.5″

The solution time for this problem was less than two seconds on an Amdahl 470V/6.

SUMMARY

This paper has described an important class of cutting stock problems that falls between the well known one and two-dimensional cases. In the example given, the two-stage cutting operation was installed in order to provide flexibility so that orders for rolls of different diameters could be combined across the width of the extruder. If a two-stage cutting operation was not used, the orders in Table 1 would have to be grouped by lineal footage and each group would then represent a separate one-dimensional cutting stock problem. It should be clear that subgrouping the orders in this way would severely restrict the patterns that could be used. Regardless of their reasons for existence, two-stage cutting stock problems do represent an important class of cutting stock problems to be solved. The example presented here, demonstrates how a modification to the Gilmore-Gomory one dimensional algorithm can be used to obtain high quality solutions to these problems.

REFERENCES

1. Eisemann K (1957) The trim problem. *Mgmt Sci.* **3**, 279–284.
2. Gilmore PC & Gomory RE (1961) A linear programming approach to the cutting-stock problem. *Ops Res.* **9**, 849–859.
3. Gilmore PC & Gomory RE (1963) A linear programming approach to the cutting-stock problem—part II. *Ops Res.* **11**, 863–888.
4. Gilmore PC & Gomory RE (1963) Multistage cutting-stock problems of two and more dimensions. *Ops Res.* **11**, 94–120.
5. Haessler RW (1976) A procedure for solving the film cutting-stock problem. ORSA-TIMS Meeting, Philadelphia.
6. Kantorovich LV (1960) Mathematical methods of organizing and planning production. *Mgmt Sci.* **6**, 366–422.
7. Metzger RW (1958) *Stock slitting, Elementary Mathematical Programming.* Ch. 8. John Wiley, New York.
8. Paul AE & Walter JR (1955) The trim problem: an application of linear programming to the manufacture of newsprint. Paper presented at annual meeting of Econometrica Society, Montreal (1954). *Econometrica* **23**(3), 336.
9. Pierce JF (1964) *Some Large Scale Production Scheduling Problems in the Paper Industry.* Prentice–Hall, Englewood Cliffs, New Jersey.

ADDRESS FOR CORRESPONDENCE: *Robert W Haessler Esq, Associate Professor of Policy and Control, The University of Michigan, Graduate School of Business Administration, Michigan* 48109, *USA.*

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 2. pp. 153 to 159
© Pergamon Press Ltd 1979. Printed in Great Britain

0306-0483 79 0401 0153$02.00/0

Inventory Control of Spares in a Transport Corporation — A Case Study

PREM VRAT
A SUBASH BABU

Indian Institute of Technology, New Delhi, India

(*Received September 1978; in revised form November 1978*)

This paper reports on a study carried out in a transport organisation to find the optimal inventory of recoverable spares in a two-level inventory system consisting of a central depot and a set of 20 depots all over the territory of the organisation, with repair facilities. The problems of the present system are diagnosed and it is observed that the upper echelon's inability to meet the lower echelon's demand has resulted in an ever-increasing shortage. A mathematical model is developed to suit the system structure and a set of optimal inventory policies for recoverable spares at both the echelons is obtained for the case studied.

INTRODUCTION

PROVIDING efficient service to commuters with limited resources is a major challenge to public transport corporations in the developing countries. Provision of sufficient spares to reduce the down-time of the vehicle is one such means of providing a good service by improved fleet availability. An inventory of recoverable spares is one such challenging problem because of the expensive nature of the items. In this paper, the policies pertaining to a recoverable spares inventory have been studied in a transport corporation of a metropolitan city in India, where a central depot caters to the needs of a number of regional depots.

There have been many attempts in the past to study the behaviour of multi-echelon inventory systems. However, the applications of this concept to a recoverable spares inventory in transport corporations have not been many. There have been some applications of the multi-echelon inventory concept for recoverable spares in the Navy and Air Force. Among various review papers on inventory systems, multi-echelon inventory systems have not found much mention. Hollier and Vrat [7] have given an extensive bibliography and observed that the actual applications have been very few. Sherbrooke [10], in one of the earliest works on multi-echelon recoverable items control, suggested a technique called METRIC using the (S-1,S) inventory policy. Muckstadt [8] extended Sherbrooke's model and proposed MOD–METRIC to permit explicit considerations of a hierarchical parts structure. Muckstadt [9] has further extended the concept to a four-echelon system and developed a model called NAVMET.

None of the models available in the literature was suitable for application to the structure of the inventory system studied in this paper. A model was first constructed to fit the environmental parameters and the same was then used to derive the optimal inventory policies for recoverable spares at the central depot as well as the various regional depots.

STRUCTURE OF THE SYSTEM

The study was conducted in a public transportation corporation with a fleet of about

TABLE 1. INFORMATION REGARDING CENTRAL WORKSHOP

No.	Stage	Rate of output (engines/day)		Time spent (days)	
		Mean	Standard deviation	Mean	Standard deviation
1	Section receiving failed engines	3.0611	3.8024	3.8721	4.3277
2	Stripping and viewing section	2.5151	2.7667	6.8108	6.4268
3	Parts stores	2.7112	2.8919	22.3652	26.9423
4	Assembling section	2.7112	2.8919	1.1001	0
5	Testing section	3.1363	1.7532	3.2603	2.5003
6	Loading dock	3.3163	2.9427	1.5904	1.6936

2000 buses operating over the entire territory of a metropolitan city catering both for urban and rural commuters. The problem was to determine the optimal number of spare bus engines to be stocked at various stocking points in the system. There is a central depot with stocking facilities and a central workshop where major engine overhauls are carried out, backed by a set of 20 subordinate (regional) depots which have storage and minor repair facilities. On an average about 100 buses are allocated to each depot to facilitate decentralised functioning of the system. The central depot is the nerve centre of the system and all purchasing of spares is done there. It also stocks a majority of spares and meets lower-level demand in the system.

Each sub-depot has a repair bay. If the failed bus engine needs minor repair, it is repaired at the sub-depot itself. For major overhauls and repairs the engine is removed from the bus and a spare engine, if available, is fitted from the sub-depot stock. If no spare is available, the demand is passed on to the central depot which meets the demand from its stock, if it has any. The failed engine is sent to the central workshop for repair and overhaul and after repair, it is deposited at the central depot in order to meet the demand of sub-depots. Inability to meet the demand either by sub-depots or the central depot results in smaller fleet availability.

In the central workshop, the engines being overhauled pass through the six stages of processing shown in Table 1.

DATA COLLECTION

Data pertaining to the number of failed engines received at the central depot, and the number of serviceable engines sent from the central depot to various sub-depots were collected over the past two years. The difference between receipt and supply is also plotted which reveals that shortages will keep on increasing on a cumulative basis due to the inability of the upper echelon to meet the demands of the lower echelon. Hence, unless something is done, it is likely that more and more vehicles will be off the road due to the shortage of spare engines, thus continually eroding the effective fleet availability. The environmental variables affecting the problem were identified and the data pertaining to the central workshop and the central and sub-depots along with the relevant cost estimates are shown in Tables 2 and 3. In order to highlight the extent of variability associated with the parameters, the standard deviation of each variable is shown; but only the actual discrete state probabilities are used in the model without approximating to any standard probability distribution function.

DEVELOPMENT OF MODEL

The structure of the inventory system and environmental parameters necessitated constructing a mathematical model to fit the system under study. The following variables are identified as relevant to the modelling of this inventory system.

Physical variables

1 Demand
2. Supply
3. Lead time
4. Spare stock (decision variables)

Vol. 7
1979

TABLE 2. RELEVANT DATA PERTAINING TO CENTRAL AND SUB-DEPOTS

No.	Type of data	Mean	Standard deviation	Notation
1	Failure rate at each depot/day	0.236	0.5697	D_j
2	No. of failed engines sent to central depot/day	0.220	0.7426	A_j
3	No. of serviceable engines received at depot/day	0.220	0.5402	B_j
4	No. of engines fitted/day at each depot	0.218	0.4792	E_j
5	Output from central workshop/day	3.316	2.7427	μ
6	Input to central workshop/ day	3.440	4.6692	λ
7	The time a failed engine spends in sub-depot in days	5.0	5.2432	L_1
8	The time a failed engine spends in central depot (workshop) while being reconditioned till it is despatched to sub-depots in days	38.89	28.7474	L_2
9	The time a serviceable engine spends in sub-depots in days	5.0	5.6212	L_3

Economic variables

1. Costs associated with failed engines at sub-depots, e.g. cost of removing/fitting a failed/ serviceable engine respectively; costs of transporting failed/serviceable engine between sub-depots and the central depot.
2. Holding costs involved at both the echelons due to the inventory of failed and serviceable engines at sub-depots and the central depot.
3. Shortage costs due to non-availability of a spare engine if demanded.
4. Cost of repair at central depot.

The objective of the model is to relate the total system cost with the spare stock at each echelon so as to find the optimal number of recoverable spares at the central and sub-depots to minimise the total system cost. The following assumptions are made:

1. The system constitutes a closed loop with no new engines being added or old ones scrapped. Since the actual additions and disposals were very insignificant, this is a valid assumption.

2. Demand of engines not met can be backlogged with resultant shortage costs due to loss

TABLE 3. ESTIMATE OF RELEVANT COST PARAMETERS

No.	Cost parameter	Denoted by	Value of the parameter
1	Cost of shortage/bus/day	s_c	Rs. 340.00
2	Factor for social cost (loss of good will)	P	0.3
3	Cost of operation/bus/day	Q	Rs. 100.00
4	Cost of replacing/engine	C_F	Rs. 40.00
5	Cost of transportation/engine/unit km	C_T	Rs. 5.00
6	Cost of Repair/engine	R_c	Rs. 2000.00
7	Cost of holding/engine/year at depot j	H_j	Rs. 5000.00
8	Cost of holding/engine/year at central depot	H_0	Rs. 5000.00
9	Cost of an engine		Rs. 35,350.00

of revenue and good will by keeping the vehicle off the road.

3. There is no lateral supply between sub-depots.

4. Shortage and other cost parameters are uniform for all the sub-depots.

5. All engines are of the same make and type.

6. Demand occurrence is independent of spare stock levels at central or sub-depots.

Notations used

n_j = number of vehicles in sub-depot j, $j = 1, \ldots, 20$

D_j = number of failures/day in depot j, a random variable with probability P_j.

A_j = number of failed engines sent to the central depot/day from depot j, a random variable with probability P'_j.

B_j = number of serviceable engines received at depot j from the central depot/day, a random variable with probability Q_j.

E_j = number of serviceable engines fitted on to vehicles/day in sub-depot j, a random variable with probability Q'_j.

N_T = total number of engines in the system.

d_{j0} = distance between jth sub-depot and the central depot.

Z = number of serviceable engines sent to the central depot from the central workshop/day, a random variable with probability $P(Z)$.

Y = number of serviceable engines issued to all sub-depots by the central depot every day, a random variable with probability $P(Y)$.

λ = average total arrival rate of failed engines at the central workshop/day.

μ = average output rate of serviceable engines from the central workshop/day.

$$= \sum_{Z=0}^{N_T} ZP(Z)$$

μ_i = output rate of ith stage of processing in the central workshop for $i = 1, \ldots, 5$.

L_1 = lead time elapsed from the instant of removal of an engine from its vehicle till it reaches the central depot, in days.

L_2 = expected time a failed engine spends at the central depot and workshop while being reconditioned till it is despatched to the sub-depot in days.

L_3 = expected time elapsed from the moment a serviceable engine is despatched from the central depot till the time it is installed in the vehicle.

S_0 = spare stock to be maintained at central depot.

S_j = spare stock to be maintained at sub-depot j.

C_F = average cost of removing and replacing an engine.

C_T = average cost of transportation of an engine to and from the central depot/unit distance.

α = expediting factor >1. This is needed if $D_j > S_j$, since it becomes necessary to expedite the transportation to avoid shortage, resulting in additional expenditure. α is taken to be 1.2 in this specific case.

H_j = average holding cost/engine/day at depot j.

H_0 = average holding cost/engine/day at central depot.

S_c = average shortage cost/backorder day.

P = penalty cost, expressed as certain percentage of S_c to include social costs of shortage.

Q = average cost of operation/bus/day.

R_c = average cost of repairing an engine.

The model

A. *Costs at sub-depot level.* The total cost at the sub-depot level includes the cost of engine failure, the cost of shortage and the cost of holding an inventory of serviceable and failed engines.

A1. *Cost associated with engine failure.* This includes the cost of removing a failed engine and fitting a serviceable engine, the cost of transportation under normal conditions as well as under emergency conditions to expedite the transfer. For all the depots the cost expression

Vol. 7
1979

for the annual planning period is as follows:

$$TCF = 365\left[\sum_{j=1}^{20}\left[\left[C_F \sum_{D_j=0}^{n_j} D_j P_j\right] + \left[C_T d_{j0} \sum_{D_j=0}^{S_j} D_j P_j\right] + \left[C_T d_{j0}\alpha \sum_{D_j=S_j+1}^{n_j} D_j P_j\right]\right]\right]. \quad (1)$$

A2. *Shortage cost.* Shortages can be caused by the sub-depot's inability to meet demand even if the central depot has a stock of serviceable spares and by the inability of the central as well as the sub-depot to meet demand. The shortage cost will be the opportunity cost of non-availability, which is loss of revenue minus operating cost, with a factor to inflate the estimate to provide for loss of good will in terms of economic repercussions. The total cost of shortage/year for the whole system is given by the following expression:

$$TCS = M[S_c(1 + P) - Q] \quad (2)$$

where

M = expected backorder days/year for the system

$$= 365\left[\sum_j\left[\sum_{D_j=0}^{n_j} P(D_j > S_j)[P(\overline{D_j - S_j} \leqslant S_0) \times [(D_j - S_j)(L_1 + L_3)] + P(\overline{D_j - S_j} > S_0) \times [(D_j - S_j)(L_1 + L_3) + (D_j - S_j - S_0)(L_2)]]\right]\right]. \quad (3)$$

A3. *Holding cost.* At sub-depot level the holding cost involves the holding cost of failed as well as serviceable engines. The expression for the total holding cost of sub-depots/year is obtained as follows:

$$THSD = 365\left[\sum_{j=1}^{20} H_j\left[\sum_{A_j=0}^{n_j}\sum_{D_j=0}^{n_j}(D_j P_j - A_j P_j') + \sum_{E_j=0}^{n_j}\sum_{B_j=0}^{n_j}(S_j + B_j Q_j - E_j Q_j')\right]\right]. \quad (4)$$

The total cost of the lower echelon can be obtained by adding equations (1), (2) and (4).

B. Costs at central depot level

The costs at the central depot level will be due to repair costs incurred in the central workshop as well as the costs of holding serviceable spare engines and an in-process-inventory of engines being overhauled in the workshop. An approximation is made here in estimating work-in-progress by treating each processing stage as a single channel queueing system with M/M/1 type queue and the total in-process-inventory obtained by adding up the inventory -(in-process) before each processing stage.

B1. *The total repairing cost/year is obtained by:*

$$TCR = 365\mu R_c. \quad (5)$$

B2. *Total holding cost due to serviceable and in-process inventory is given by:*

$$TCHC = 365\left[H_0\left[S_0 + \mu - \sum_{Y=0}^{N_T} YP(Y)\right] + H_0\left[\sum_{i=1}^{5}\frac{\lambda}{\mu_i - \lambda}\right]\right]. \quad (6)$$

The total system cost for both the echelons is obtained by adding equations (1), (2), (4), (5) and (6). The decision variables S_0 and S_j are to be determined so as to minimise the total system cost. It can be seen that 365 is a common multiplier in all the constituent equations of total system cost and hence minimising the total system cost/year is an objective equivalent to minimisation of the total system cost/day.

ANALYSIS OF RESULTS

The model developed for optimization of the recoverable spares inventory at both the echelons was used to find the optimal policies for the case under discussion. Values of parameters obtained through Tables 1–3 were substituted in the total system cost equation. Since no analytical optimization technique was available to suit the model, the numerical method of optimization by computer-search was tried. S_j was varied between 0 and 4 as it was observed from data that D_j never takes a value greater than 4. S_0 was varied from 1 to 19. For each combination of S_0 and S_j, all the costs (A1, A2, A3, B1 and B2) were obtained as well as the total system cost/year. These values are shown in Table 4.

The results indicate that the model behaves in a plausible manner and all individual cost

TABLE 4. TOTAL SYSTEM COST AS FUNCTION OF S_0 AND S_j

S_j \ S_0	1	2	3	4	5	6	7	8	9	10
0	17116472	11345722	8439862	8445262	8450602	8456062	8461462	8466862	8472262	8277662
1	6741851	5289983	5295383	5300783	5306183	5311583	5316983	5322383	5327783	5333183
2	4490671	4496070	4501471	4506871	4512271	4517671	4523071	4528471	4533871	4539271
3	4382277	4387677	4393077	4398477	4403877	4409277	4414677	4420077	4425477	4430877
4	4492805	4498205	4503605	4509005	4514405	4519805	4525205	4530605	4536005	4541405

S_j \ S_0	11	12	13	14	15	16	17	18	19
0	8483062	8488462	8493862	8499262	8504662	8510062	8515462	8520862	8526262
1	5338583	5343983	5349383	5354783	5360183	5365583	5370983	5376383	5381783
2	4544671	4550071	4555471	4560871	4566271	4571671	4577071	4582471	4587871
3	4436277	4441677	4447077	4452477	4457877	4463277	4468677	4474077	4479477
4	4546805	4552205	4557605	4563005	4568405	4573805	4579205	4584605	4590005

components have a behaviour pattern, which is even intuitively consistent. As S_0 or S_j or both are increased, the shortage cost is reduced and the holding cost increased, as expected. The results in Table 4 indicate that the total system cost is more sensitive to S_j, than to S_0. The economic behaviour of the system is hypersensitive (for the case under discussion) when S_0 is changed from 1 to 2 and 2 to 3 and S_j changed from 0 to 1 and 1 to 2. The global minimum cost is obtained at $S_0 = 1$ and $S_j = 3$. For $S_j = 3$, any increase in S_0 will not result in any appreciable increase in the total system cost, indicating the insensitivity of the system to S_0, at optimal S_j. There is a marginal difference between the total system cost at $S_j = 2$ and $S_j = 3$. Since reduction in the total inventory works out to 20 engines, indicating an investment reduction of Rs. 700,000, this sub-optimal decision ($S_j = 2$) might be preferable as the problem of initial investment is appreciable to the organisation studied. The present practice was not to have any spare engine at either level, and obviously it was creating problems of acute shortage as was evident from the initial diagnostic analysis reported in preceding sections.

CONCLUSION

The model developed in this paper, though specific to this situation, can be used elsewhere for recoverable spares inventory control with little modification depending upon the system structure. In the context of the transportation corporation studied, the model provided a potential for cost reduction if inventory policies as prescribed by the model are implemented. It is also possible to study the effect of increased maintenance facilities (and consequent reduction in throughput time) on the spare level. Further, it is possible to investigate the effect of location of central/sub-depots on stock levels, *etc.*

However, the model could possibly be further refined by relaxing the assumptions that there is no lateral supply between sub-depots and that there are no additions and condemnation of engines in the system. Permitting lateral supplies would require a very efficient information system. In the model the value of S_j is taken to be same for each sub-depot, to simplify computational effort. The model can be extended to a three-echelon system and work is in progress in that direction. Even in its present form the model offers a scope for substantial cost reduction.

Vol. 7
1979

ACKNOWLEDGEMENT

The authors wish to thank the referee for the constructive suggestions which have improved the presentation of this paper.

REFERENCES

1. AGGARWAL SC (1974) A review of current inventory theory and its applications, *Int. Jl. Prod. Res.* **12**(4), 443.
2. ALLEN SG & D'ESOPO DA (1968) An ordering policy for repairable stock items. *Ops. Res.* **16**(3), 669.
3. CLARK AJ (1972) An informal survey of multi-echelon inventory theory. *Nav. Res. Log. Q.* **19**(4), 621.

4. FEENEY GJ & SHERBROOKE CC (1966) The (S-1,S) Inventory policy under compound poison demand. *Mgmt. Sci.* **12**(5), 391.
5. HABER SE & SITGREAVES R (1975) An optimal inventory model for the intermediate echelon when repair is possible. *Mgmt. Sci.* **21**(6), 638.
6. HILLIER F & LIEBERMAN G (1967) *Introduction To Operations Research.* Holdenday, San Francisco.
7. HOLLIER RH & PREM VRAT (1976) A Review of the multi-echelon inventory control research and applications, Technical Report, University of Birmingham, Engl.
8. MUCKSTADT JA (1973) A model for a multi-item, multi-echelon, multi-indenture inventory system. (MOD-METRIC). *Mgmt. Sci.* **20**(4), 472.
9. MUCKSTADT JA (1976) A Four echelon model for determining the optimal quantity and distribution of Navy Spare Aircraft engines. (NAVMET). Technical Report No. 263, School of O.R. Cornell University.
10. SHERBROOKE CC (1968) A Multi-echelon technique for recoverable item control, (METRIC). *Ops Res.* **16**, 122.
11. SMITH SA (1977) Optimal inventories for a (S-1,S) system with no back orders. *Mgmt. Sci.* **23**(5), 522.

ADDRESS FOR CORRESPONDENCE: *Dr Prem Vrat, Department of Mechanical Engineering, Indian Institute of Technology, New Delhi* 110 029, *India.*

ol. 7
979

BLANK
PAGE

Vol. 7
1979

MEMORANDA

Vol. 7
1979

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 2. pp. 163 to 168
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0401-0163$02.00/0

Four Examples of Artless Modelling

INTRODUCTION

MOST ACADEMIC attention in Management Science and Operational Research has been devoted to the more algorithmic and mathematically deep aspects of the subject which readily lend themselves to systematisation. In comparison the 'art' of model building has been rather neglected. There is often a large artistic component in the building of models. This is very difficult to teach although it can be demonstrated and, to some extent, acquired by (good) experience. The neglect of the modelling aspect sometimes results in quite surprisingly (to the author) badly built models or oversights. Technicalities in the modelling and solution procedures cause a blindness which results in a failure to 'see the wood for the trees'. Four examples are presented here, three from the author's experience and one from that of another practitioner. It is hoped these will demonstrate the need for common sense thinking when modelling.

EXAMPLE 1: BLINDNESS IN MODELLING

The problem here involved blending a number of different grades of mineral together in order to achieve certain quality stipulations at minimum costs. As such it was a classic application for Linear Programming. The problem arose each week. Each week's problem gave rise to a minute (6 constraint) LP model. The problem was, however, complicated by its dynamic nature. The ingredient mineral need not be entirely used nor the blended mineral entirely sold in the weeks in which it was extracted. This possibility of passing on minerals between successive weeks was dealt with by 'recursively' solving the LP models for each week. By means of a FORTRAN program interfaced to the LP package used the surplus of one week's solution was fed into the next week's model before it in turn was solved. This procedure was repeated (automatically) for all 52 weeks in the year. The year's solution was then used to adjust the starting conditons for a repeat of the whole procedure. Carrying out this cycle a number of times an acceptable solution was obtained. There appeared (to the author) no guarantee that the process would result in an optimal solution. The method did, however, work and produce reasonable results.

What the model builders had failed to realise was that the dynamic aspect of their problem could also be modelled using LP. It would have been reasonably straight forward to introduce variables as follows:

u_{it} = Quantity of ingredient i produced in week t
v_{it} = Quantity of ingredient i blended in week t
w_{it} = Quantity of ingredient i passed on at end of week t
x_t = Quantity of blended mineral produced in week t
y_t = Quantity of blended mineral sold in week t
z_t = Quantity of blended mineral passed on at end of week t

The following linear relationships could then have been stipulated.

$$w_{it-1} + u_{it} - v_{it} - w_{it} = 0$$
$$z_{t-1} + x_t - y_t - z_t = 0.$$

These constraints could then have been incorporated in a *multiperiod* LP model in addition to all the 52 weekly blending models. Such a model would have been much larger than the original LP models but still of only a moderate size in comparison to what can be solved by modern computer systems. Although such a model was not built it would almost certainly have been quicker to solve than the original, would have avoided the writing of a special purpose routine and could have been guaranteed to produce the optimum solution.

In this case there was obviously a failure to think beyond the simple type of LP constraints to the possibility of modelling dynamic relationships such as those above. The modelling of such structured situations is described [3].

EXAMPLE 2: INELEGANT MODELLING

This problem involved making production and inventory decisions over a number of successive months. The problem had a rather similar structure to that of Example 1 but a multiperiod LP model was in this case built. Variables were introduced with the following interpretations:

x_{it} = Quantity of product i produced in month t.

There were known demands d_{it} for product i in month t.

All the productive capacity constraints were modelled for each month. In the objective function both production and storage costs were represented for each month.

The possibility of under or over-producing in certain months and coping with shortfall or excess by storing (at a cost) was modelled by the following relationships (for all i):

Month 1 $x_{i1} = d_{i1}$

Month 2 $x_{i2} = d_{i2} - x_{i1} + d_{i1}$

Month 3 $x_{i3} = d_{i3} - x_{i2} + d_{i2} - x_{i1} + d_{i1}$

Month 4 $x_{i4} = d_{i4} - x_{i3} + d_{i3} - x_{i2} + d_{i2} - x_{i1} + d_{i1}$

etc.

Clearly these relationships get longer and longer as the months proceed. This model was not incorrect, but most practitioners would regard it as inelegant. The model builder has obviously related all quantities back to the first month.

By introducing extra variables (although this makes the model larger) the relationships can be thought of 'recursively' linking together only immediately successive months. The following extra variables are introduced:

y_{it} = Quantity of product i stored from month t to

month $t + 1$. The above relationships can then be remodelled (for all i) as:

Month 1 $\quad x_{i1} - y_{i1} = d_{i1}$

Month 2, 3, .. $\quad y_{it-1} + x_{it} - y_{it} = d_{it}$

EXAMPLE 3: INCORRECT MODELLING

This example was described to the author by Müller-Merbach [2], who suggests a systematic notation for representing the type of problem described here. We purposely do not follow his notation in order to demonstrate how easy it is to model incorrectly without such systematisation. The problems concern the flow of materials through a network of processes as found in, for example, oil refineries. Two examples are given here.

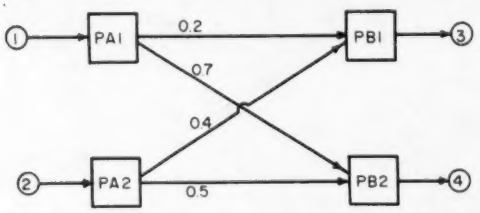


FIG. 1. *Flow through a network.*

Figure 1 illustrates the first example of flows into and out of two types of processes. Process A (PA1 and PA2) *splits* the input material in *fixed* proportions into two types of output material. (e.g. a distillation process splits crude oils into different grades of petrol). Process *B* (*PB*1 and *PB*2) *mixes* its inputs together in *flexible* proportions to produce an output (e.g. the blending together of ingredients). In this example:

PA1 splits each unit of (1) into 0.2 units and 0.7 units respectively of two intermediate materials.

PA2 splits each unit of (2) into 0.4 units and 0.5 units respectively of two intermediate materials.

PB1 and PB2 mix the intermediate materials together (with no loss of mass) as indicated in Fig. 1 to produce final materials (3) and (4).

If the quantities of (1), (2), (3) and (4) involved are represented by variables x_1, x_2, x_3 and x_4 the building of a model demands the representation of the relationships between these quantities. The correct formulation is:

$$x_3 = 0.2x_1 + 0.4x_2$$

$$x_4 = 0.7x_1 + 0.5x_2.$$

The second example is illustrated by Fig. 2.

Process *C* (*PC*1 and *PC*2) is a *flexible* conversion process which may convert a unit of material into a unit of one or other materials (e.g. milk into butter or milk into cheese). Clearly some of the input may be used to produce one material (butter) and the rest to produce the other material (cheese). Process *D* (*PD*1 and *PD*2) takes inputs in *fixed* proportions and produces an output material (e.g. the production of certain fuel oils in a refinery requires the blending of different oils in fixed proportions). In this example:

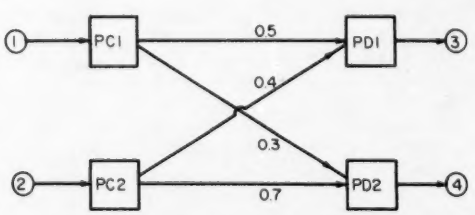


FIG. 2. *A second example of network flow.*

PC1 converts each unit of (1) to a unit of one or other of two intermediate outputs (or a mixture of the two).

PC2 converts each unit of (2) in a similar fashion.

PD1 takes its two inputs and produces its output of (3) in the fixed ratio 0.5:0.4:1.

PD2 takes its two inputs and produces its output (4) in the fixed ratio 0.3:0.7:1.

The quantities of (1), (2), (3) and (4) will again be represented by variables x_1, x_2, x_3 and x_4. It is tempting to formulate the relationships between them (analogously to the former example) as:

$$x_3 = 0.5x_1 + 0.4x_2$$

$$x_4 = 0.3x_1 + 0.7x_2.$$

A little reflection shows this must be wrong (for example the formulation would allow positive quantities of (3) and (4) to be produced using none of (2) breaking the fixed proportion requirement of process *D*). The correct formulation is:

$$x_1 = 0.5x_3 + 0.3x_4$$

$$x_2 = 0.4x_3 + 0.7x_4.$$

Müller-Merbach has experience of a model built making this mistake many times over thus producing a wildly invalid representation of the real life problem.

This example illustrates the danger of failing to analyse a simple situation where an apparent analogy presents itself.

EXAMPLE 4: BLINDNESS IN RECOGNISING A SOLUTION

This example concerned a multiperiod planning problem stretching over 20 years. Many investment decisions had to be made during this time. These decisions were interrelated. Investment in a piece of plant at any time affected both investments in other pieces of plant at that same time and investments in the same piece of plant at future times. This complex problem was formulated as a multiperiod Integer Programming model.

To start with a small four period model was built in which each of the periods was an aggregated 5 year time span. The solution produced obviously allowed no changes in investment plans within any of the 5 year periods. A disaggregated model was then constructed consisting of 20 one year time periods. This model was obviously much larger and more difficult to solve. It was run for many hours on a computer and not even a feasible solution was found (let alone an optimal one).

The performance of most mathematical programming algorithms can be greatly improved by providing an initial feasible solution. For Integer Programming models the commonly used Branch and Bound algorithm is no exception. A feasible solution provides a

Vol. 7
1979

bound on the value of the objective function which can be used to good effect in the tree search procedure.

In this example, a feasible solution was readily to hand if only the solver (who was also the modeller) had realised it. The solution to the aggregated four period model should obviously have suggested a feasible solution to the 20 period model in which variables remained unchanged within each span of 5 years. This observation would (in the author's opinion) have saved much computer time and money. Moreover, the observation requires no technical but rather an ability to step back and think in common sense terms about the relationship between the models and the situation represented.

CONCLUSIONS

Modelling is often given insufficient attention in Management Science. The potential of Linear Programming to model other than simple situations is not as widely known as it might be [3]. What may seem trivial to many practitioners is far from obvious to those less experienced.

The use of an established Management Science approach such as Linear Programming can easily cause people to lose their original or critical faculties and apply the approach unthinkingly. Model building is to some extent an art in a rather analogous way to computer programming as discussed in Knuth [1].

Although the examples described here relate to Mathematical Programming models similar conclusions are likely to apply more widely.

REFERENCES

1. KNUTH DE (1974) Computer programming as an art. *Jl ACM*, **17**, pp 667–673.
2. MULLER-MERBACH H (1978) Entwurf von input-output-modellen. *Proc. Ops Res.*, **7**, pp 521–532 (ed. K Brockhoff *et al*). Physica-Berlag, Wurzburg, Wien.
3. WILLIAMS HP (1978) *Model Building in Mathematical Programming*. John Wiley, London.

Professor HP Williams
(*November* 1978)

Department of Business Studies
University of Edinburgh
50 George Square
Edinburgh

Vol. 7
1979

Some Useful Ratios in Evaluating Performance

ABSOLUTE values of key variables in an industrial enterprise are rarely adequate to tell us how the enterprise performs. For example, measures of output, sales volume or number of employees are fairly good indicators of the size of the operation, but for performance evaluation purposes we need to know how these measures change with time, what resources are needed to produce certain outputs, and how the resources mix compares with other operations. The use of ratios, which link inputs and outputs or express relationships between different measures of inputs or outputs, has therefore become common practice among business analysts. The number of possible ratios that can be considered is legion (see, for example, Westwick [5]), and long lists are often constructed in a vain attempt to cater for every conceivable eventuality, although in practice managers and analysts are very selective in what they use, and sometimes they even devise their own measures to suit the environment in which they operate. The purpose of this note is to present a list of manageable proportions in the hope that it is adequate to cover most of the contingencies encountered in practice and thereby provide a range from which a suitable choice can be made in particular instances.

Ratios can be divided into two groups, the first relates to measures of performance at any given time (seven categories are listed below for relationships between inputs and outputs in various combinations, as well as ratios reflecting financial, market and productivity considerations), and the second indicates how measures of change can be handled by the method of incremental analysis.

NOTATION

Inputs

I_F = fixed assets (machinery, buildings, land)
I_W = working capital (current assets minus current liabilities)
$I = I_F + I_W$ = capital employed
N = number of employees $= N_1 + N_2$
N_1 = direct labour
N_2 = indirect labour
W = wages and salaries $= W_1 + W_2$
W_1 = cost of the N_1 employees
W_2 = cost of the N_2 employees
M = materials $= M_1 + M_2$
M_1 = cost of raw materials
M_2 = cost of bought components
CC = capital charges
OE = other expenses

Intermediates

H = manhours
Q = stocks (in monetary value)
C = total cost $= W + M + CC + OE$
p = unit price $= R/V$
c = unit cost $= C/V$
$p - c$ = unit profit

Outputs

V = production volume (in physical units; for further details of how to measure output see [2])
S = sales (in physical units; it is often assumed that $V = S$, though such an assumption is not always valid)
R = sales revenue $= pS$ (assumed often to be pV)
Z = profit (usually before tax) $= \bar{C} - \bar{R} = (p - c)V$

MEASURES OF PERFORMANCE

1. *Relationships between inputs*

I_F/I = ratio of fixed assets for future investment
I_F/N = fixed assets per employee
I/N or I/W = total assets per employee or per cost of employees
N_1/N (or W_1/W) or N_2/N (or W_2/W) = direct labour ratio
M_1/M or M_2/M = bought components indicator

2. *Relationships between intermediate and inputs or outputs*

$C/V = c$ = unit cost
$\alpha = (p - c)/p$ = profit margin
H/N_1 = hours per direct employee (or W_1/H = average hourly cost of direct labour)
H/V = manhour per unit of output
Q/I or Q/I_W = indicator of stock holding
R/Q = stock turnover ratio
M/C = proportion of material cost
W/C = proportion of labour cost
CC/C = proportion of capital cost
OE/C = proportion of other expenses

Note: $\dfrac{M}{C} + \dfrac{W}{C} + \dfrac{CC}{C} + \dfrac{OE}{C} = 1$

3. *Relationships between outputs*

$Z/R = \alpha$ = profit margin (as above)
$R/V = p$ = unit price
Z/V = unit profit $= p - c$ (as above)

4. *Relationships between outputs and inputs*

Input	Production (or sales revenue)	Profit
I	V/I (or R/I)	$Z/I = r$ = rate of return on investment
I_F	V/I_F (or R/I_F)	Z/I_F
N	V/N (or R/N)	Z/N

M/V = unit material cost
W/V = unit labour cost

CC/V = capital charges per unit
OE/V = other expenses per unit

Note: $\frac{M}{V} + \frac{W}{V} + \frac{CC}{V} + \frac{OE}{V} = c$

5. *Other financial ratios*

$$\text{current ratio} = \frac{\text{current assets}}{\text{current liabilities}}$$

$$\text{liquid ratio ('acid test')} = \frac{\text{cash + debtors}}{\text{current liabilities}}$$

$$\text{assets ratio} = \frac{\text{current assets}}{\text{fixed assets}}$$

$$\text{debtors credit} = \frac{\text{debtors}}{\text{av. daily sales}}$$

$$\text{debt ratio} = \frac{\text{total debt}}{\text{total assets}}$$

$$\text{gearing ('leverage')} = \frac{\text{preference capital + long term loans}}{\text{equity capital}}$$

$$\text{'times cover'} = \frac{\text{profit after depreciation}}{\text{preference dividend + long term loan interest}}$$

6. *Market indicators*

market share
share price
eps (earnings per share)

$$\frac{eps}{\text{share price}}$$

P/E (share price/earnings)

$$\frac{\text{share price}}{\text{Financial Times or Dow Jones Index}}$$

$$\frac{\text{share price}}{\text{sector share index}}$$

7. *Added value ratios*

AV = added value
= sales income − purchase expenditure
$= R - M = W + CC + OE + Z$

A typical argument of proponents of added value measures runs as follows: "Once capital has been invested, the depreciation, loan interest and profit on the increased investment all have to be earned from added value. So the added value per £ fixed asset and per £ capital employed are important—they should remain constant" [4]. It is not clear why they should, but the performance ratios most commonly mentioned are:

AV/W
AV/N
AV/I_W
AV/I_F
AV/I
AV/day
AV/sq. ft. (or AV/sq. m.)
AV/fuel costs

8. *Productivity ratios* (for further details see [2])

labour productivity = V/N (or V/H) = output/man (or manhour)
materials productivity = V/M = output/materials
capital productivity = output/capacity

INCREMENTAL ANALYSIS

More often than not, managers are interested in the degree to which certain measures undergo a change. For example, if the unit cost c in one time period is subjected to an incremental change δc in a subsequent period, then attention is focused on $c^* = \delta c/c$ which measures the relative change in c, and similarly Z^* is the ratio of the change in profit to the original profit. The relative change in any of the variables or ratios listed above may therefore become the subject of managerial scrutiny.

In general terms it can be shown [1] that if $A = ab$ and if the variable a in one time period changes to $a(1 + a^*)$ in a subsequent period (where $a^* = \delta a/a$) and similarly if the variable b changes to $b(1 + b^*)$ then the relative change $A^*(= \delta A/A)$ is $A^* = a^* + b^* + a^*b^*$. Similarly, if $A = abc$, then $A^* = a^* + b^* + c^* + a^*b^* + b^*c^* + c^*a^* + a^*b^*c^*$

Also, if $A = a + b$ then it can be shown that

$$A^* = a^* \frac{a}{A} + b^* \frac{b}{B}$$

For the case $C = W + M + CC$ (for the sake of simplicity EO is ignored here) it follows that

$$C^* = W^*\left(\frac{W}{C}\right) + M^*\left(\frac{M}{C}\right) + CC^*\left(\frac{CC}{C}\right)$$

Similarly, for the unit cost:

$$c^* = \left(\frac{C}{V}\right)^* = \underbrace{\left(\frac{W}{V}\right)^* \frac{W}{C}}_{= w^*} + \underbrace{\left(\frac{M}{V}\right)^* \frac{M}{C}}_{= m^*} + \underbrace{\left(\frac{CC}{V}\right)^* \frac{CC}{C}}_{= cc^*}$$

EXAMPLE

1. *Total cost*

	C	W	M	CC	V
base year 1	100 =	20 +	60 +	20	100
year 2	120 =	30 +	70 +	20	110
contribution to change	20 =	10 +	10 +	0	10
% change	20	50	16.7	0	10
cost proportions		0.2	0.6	0.2	total = 1
contribution		10	10	0	total = 20
% contribution	$\frac{10}{20} \times 100 =$	50	50	0	total = 100

2. *Unit cost*

	c		$\frac{W}{V}$		$\frac{M}{V}$		$\frac{CC}{V}$
base year 1	100	=	20	+	60	+	20
year 2	109.1	=	27.3	+	63.6	+	18.2
% change	9.1		36.5		6		−9
% change	9.1		7.3	+	+3.6	−	1.8
			w^*		m^*		cc^*

Change in unit cost [3]

If

$$c = \text{unit cost} = s + \frac{F}{V}$$

where

s = variable cost/unit
F = fixed cost
V = volume

then

$$c^* = f_1 s^* + \frac{f_2}{1 + V^*}(F^* - V^*)$$

where

$f_1 = s/c$ = proportion of the variable cost in the total unit cost
$f_2 = (F/V)/C$ = proportion of the fixed costs in the total unit cost

and

$$f_1 + f_2 = 1$$

Change in rate of return

If rate of return

$$r = \frac{Z}{I} = \frac{(p - c)V}{I}$$

then approximately

$$r^* = \frac{p^*}{\alpha} - \frac{1 - \alpha}{\alpha} c^* + V^*$$

REFERENCES

1. Eilon S (1975) Changes in profitability components. *Omega* **3**(3), 353–354.
2. Eilon S, Gold B & Soesan J (1976) *Applied Productivity Analysis for Industry.* Pergamon Press. Oxford.
3. Eilon S & Phocas-Cosmetatos G (1976) *Analysis of profitability components.* Proc. EURO II (Second Eur. Congr. Op. Res.) 143–147.
4. Smith G (1976) Understanding productivity. *In. Comml Train.* **8**, 485–489.
5. Westwick CA (1973) *How to Use Management Ratios.* Gower Press.

Samuel Eilon

Department of Management Science
Imperial College
London, SW7 2BX
UK

(*November* 1978)

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 2. pp. 169–172
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0401-0169$02.00/0

International Register of Theses and Dissertations

The following are obtainable at the MIT Microreproduction Laboratory, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge MA 02139, USA:

Alvarez-Morodo JA (1974) A Pseudo-Lexicographic Model for Consumer Brand Reference. SM Thesis, MIT, Cambridge, MA.

Angelicoussis JA (1974) Shipleasing as a Prospective Method of L/T Financing for International Shipowners. SM Thesis, MIT, Cambridge, MA.

Aylward TJ (1974) The Internal Corporate Venture—Its Present Status as a Corporate Growth Strategy. SM Thesis, MIT, Cambridge, MA.

Barker KJ (1974) Attitudes of Foremen in the Automobile Industry. SM Thesis, MIT, Cambridge, MA.

Berry CJ (1974) Energy Management: Energy Conservation and the Environmental Conflict. SM Thesis, MIT, Cambridge, MA.

Borzutzky R (1974) Control of Technology in Latin America: A Study of Multinational Corporations Perception. SM Thesis, MIT, Cambridge, MA.

Cannistraro RS (1974) The Predictive Ability of Price/Earnings Ratios for Determining Risk–Adjusted Performance of Common Stocks. SM Thesis, MIT, Cambridge, MA.

Chigusa T (with **Meyer W**) (1974) Short Term Price Behavior on the Commodities Futures Market. SM Thesis, MIT, Cambridge, MA.

Clements AJ (with **Van Assche B & De Crouy-Chanel H**) (1974) Study of a London Merchant Bank. SM Thesis, MIT, Cambridge, MA.

Cohen JJ & Cook WE (1974) Integrated Production Planning and Control in an Upholstery Textile Firm. SM Thesis, MIT, Cambridge, MA.

Cooper RE (1974) Japanese and US Market Requirements for Pure Water. SM Thesis, MIT, Cambridge, MA.

Deren DJ (1974) The Decision to Leave Stock Certificates with the Broker: An Examination of the Investor's Decision Process. SM Thesis, MIT, Cambridge, MA.

Dyer JC (1974) Financial Leverage, Business Risk, and the Tax Subsidy Due to Debt: An Empirical Study. SM Thesis, MIT, Cambridge, MA.

Eirich PL (1974) An Information System Design Analysis for a Research Organization. SM Thesis, MIT, Cambridge, MA.

Fifield JJ (1974) Financial Institutions and Economic Development Strategies for the United States of America. SM Thesis, MIT, Cambridge, MA.

Folk RE (with **Robertson JL**) (1974) A System Dynamics Representation of Monetary and Fiscal Policy. SM Thesis, MIT, Cambridge, MA.

Fortson RM (1974) Application of the Sequential Unconstrained Minimization Technique in a Systematic Containership Design. SM Thesis, MIT, Cambridge, MA.

Fraile A (1974) Western Perceptions of the Manager in the People's Republic of China. SM Thesis, MIT, Cambridge, MA.

Frank R (1974) Publicly Aided Housing in Germany: Current Practices and Their Relevance to United States Housing Policies. SM Thesis, MIT, Cambridge, MA.

Freitas HM (with **Freitas JS**) (1974) Integrating Organizational Objectives with Individual Goals in a R&D Center—An Action-Research Approach. SM Thesis, MIT, Cambridge, MA.

Gruber JM (1974) A Model for Product Selection and Machine Assignment in a Job Shop. SM Thesis, MIT, Cambridge, MA.

Halperin JA (1974) The Relationship of Background and Employment Factors to Theory X–Theory Y Assumptions and Managerial Style. SM Thesis, MIT, Cambridge, MA.

Hamilton J (1974) Aspects of Foreign Direct Investment in US Real Property. SM Thesis, MIT, Cambridge, MA.

Hendrickson GP (1974) Potential Corporate Services Provided Using Terminals Linked to Bank Computers. SM Thesis, MIT, Cambridge, MA.

Hoberman A (with **Meyer W**) (1974) An Econometric Analysis of Real Estate Capitalization Rates. SM Thesis, MIT, Cambridge, MA.

Hsu CPY (with **Tso DN**) (1974) Analysis of US–Canadian Forward Exchange Market. SM Thesis, MIT, Cambridge, MA.

Inman EH (with **Tchen S**) (1974) A Formal Approach to Risk Analysis in Real Estate. SM Thesis, MIT, Cambridge, MA.

Janzen CH (with **Koval RJ**) (1974) Salesmen Compensation and Motivation. SM Thesis, MIT, Cambridge, MA.

Kamiryo H (1974) A Comparison of Financial Objectives and Behavior in Japanese and American Firms. SM Thesis, MIT, Cambridge, MA.

Kanner SR (1974) The Impact of Cost Reimbursement on the Financing of Voluntary Hospitals. SM Thesis, MIT, Cambridge, MA.

Karash R (1974) Marketing in the High Fidelity Industry. SM Thesis, MIT, Cambridge, MA.

Kingery JC (1974) Lease Versus Buy: An Application to Airline Equipment Acquisition. SM Thesis, MIT, Cambridge, MA.

Kotecha M (1974) Assessment of the Impact of Alternative Federal Funding Mechanisms on the National Subsidized Family Planning Services Programs. SM Thesis, MIT, Cambridge, MA.

Krygier R (with **Barker KJ**) (1974) Attitudes of Foremen in the Automobile Industry. SM Thesis, MIT, Cambridge, MA.

Kunishige A (1974) On Determining an 'Ideal' Decision Support System. SM Thesis, MIT, Cambridge, MA.

Lantos PS (1974) Demand Analysis and Need Measurement for Planning Mental Health Facilities. SM Thesis, MIT, Cambridge, MA.

Lee WWL (1977) When Values Conflict: An Application of Multiattribute Decision Analysis to Marine Sand and Gravel Mining. ScD Thesis, MIT, Cambridge, MA.

Lim JS (1974) Strategic Planning in Bank Holding Companies: Theory and Practice. SM Thesis, MIT, Cambridge, MA.

Lincoln JD (1974) Relevant Costing System: A Conceptual Scheme for Cost Analysis in the Public Sector. SM Thesis, MIT, Cambridge, MA.

Loeb SM (1974) Design of a Visual Interface for Relational Data Base Systems. SM Thesis, MIT, Cambridge, MA.

Ludwig RS (1974) Forecasting Short-Term Savings Deposit Flows: An Application of Time Series Models and a Regional Analysis. SM Thesis, MIT, Cambridge, MA.

Madrazo A (1974) Personal and Environmental Factors Affecting Engineer Career Development. SM Thesis, MIT, Cambridge, MA.

Madson T (1974) Industrial Democracy: A Human Needs Approach. SM Thesis, MIT, Cambridge, MA.

Marchand CP (1974) An Inquiry into the Use of Some Management Techniques in Various Sectors of the American Industry. SM Thesis, MIT, Cambridge, MA.

Massey CL (1974) Decision Making in Human Resource Management. SM Thesis, MIT, Cambridge, MA.

McLellan DR (1974) The Automobile Industry and Government. SM Thesis, MIT, Cambridge, MA.

Mendes Filho NT (with **De Moraes CAC**) (1974) Reciprocal Colleague Roles in a R&D Organization. SM Thesis, MIT, Cambridge, MA.

Merrigan JA (1974) Status and Potentiality of Photovoltaic Energy Conversion. SM Thesis, MIT, Cambridge, MA.

Merritt JP (1974) A Demand Forecasting Model for a Machine Tool. SM Thesis, MIT, Cambridge, MA.

Metz RW (with **Wexler D**) (1974) The Pricing of Delayed Floating Convertible Bonds. SM Thesis, MIT, Cambridge, MA.

Montaniel MB (1974) A Cost-Effectiveness Analysis of the Expanded-Function Dental Auxiliary Concept. SM Thesis, MIT, Cambridge, MA.

Moser RB (1974) Turnover and Labor Market Organization. SM Thesis, MIT, Cambridge, MA.

Murphy KW (1974) A Profile of the Effective Industrial Product Management—Electric Industry Product Manager. SM Thesis, MIT, Cambridge, MA.

Ngai SYA (1974) Long-Distance Commuting as a Solution to Geographical Limitation to Career Choices of Two-Career Families. SM Thesis, MIT, Cambridge, MA.

Njomo EN (1974) The Effects of Dealer Financing on the Costs of Mobile Home Ownership. SM Thesis, MIT, Cambridge, MA.

Olagundoye OB (with **Owusu P**) (1974) Management Control, Development and the Multinational Corporation. SM Thesis, MIT, Cambridge, MA.

Patzelt L (with **Holzer D**) (1974) Computer Services for Business Schools: A Procedure for Analysis and Application to the Sloan School of Management. SM Thesis, MIT, Cambridge, MA.

Powell TC (1974) A Survey of Consumer Attitudes Concerning Automotive Regulatory Laws. SM Thesis, MIT, Cambridge, MA.

Poza Valle EJ (1974) Career Paths in Organization Development: The Practitioner and his Environment. SM Thesis, MIT, Cambridge, MA.

Vol. 7
1979

RANFT EL (1974) Foreign Wholly Owned Subsidiaries: How Independent are they? SM Thesis, MIT, Cambridge, MA.

RIORDAN E (1974) The Impact of Demographic Profile Shifts on the Social Security System. SM Thesis, MIT, Cambridge, MA.

ROANE AF (1974) Organization Dynamics: A System Dynamics Study of the Behavior of the Firm in a Changing Environment. SM Thesis, MIT, Cambridge, MA.

ROBSON ES (with **WYNNE R**) (1974) Change Strategies: A Survey and Analysis. SM Thesis, MIT, Cambridge, MA.

ROCHARD JEA (1974) Management of Urban Land Banks. SM Thesis, MIT, Cambridge, MA.

SAULNIER YRM (1974) Growth as a Corporate Objective and Some of its Financial Implications. SM Thesis, MIT, Cambridge, MA.

SCHULMEYER GH (1974) Preferences for Information Sources in a Marketing Decision Process. SM Thesis, MIT, Cambridge, MA.

SIT WH (1974) Energy Management in Supermarkets. SM Thesis, MIT, Cambridge, MA.

SMITH DH (with **COGGINS PE**) (1974) The Manufacturing Foreman Today in the Petroleum Industry: His Role and His Problems. SM Thesis, MIT, Cambridge, MA.

SUOJANEN WW (1974) A Case Study of the Longitudinal Effects of a Job Enrichment Program. SM Thesis, MIT, Cambridge, MA.

VON BRAUN C-F (1974) The Technological Gatekeeper—An International Phenomenon? SM Thesis, MIT, Cambridge, MA.

WALLACE PR (1974) Marketing Strategies for Radically New Consumer Concepts. SM Thesis, MIT, Cambridge, MA.

ALTER SL (1975) A Study of Computer Aided Decision Making in Organizations. PhD Thesis, MIT, Cambridge, MA.

ANDERSON DF (1977) Mathematical Models and Decision Making in Bureaucracies: A Case Story Told from Three Points of View. PhD Thesis, MIT, Cambridge, MA.

BEHRENS W (1975) A Policy Structure for Monetary Control Decisions. PhD Thesis, MIT, Cambridge, MA.

BERNHART MH (1977) Model Clinics: A Medium for Disseminating Administrative Innovations. PhD Thesis, MIT, Cambridge, MA.

BITRAN GR (1975) Admissible Points and Vector Optimization: A Unified Approach. PhD Thesis, MIT, Cambridge, MA.

BODILY SE (1976) Collective Choice with Multidimensional Consequences. PhD Thesis, MIT, Cambridge, MA.

BROOKS RE (1975) Allocation of Natural Gas in Times of Shortage: A Mathematical Programming Model of the Production, Transmission, and Demand for Natural Gas under Federal Power Commission Regulations. PhD Thesis, MIT, Cambridge, MA.

BUMSTEAD DC (1975) Organization Structure and Performance: An Evaluative Study of Management in a Residential Institution. PhD Thesis, MIT, Cambridge, MA.

CANDEA D (1977) Issues of Hierarchical Planning in Multi-Stage Production Systems. PhD Thesis, MIT, Cambridge, MA.

CHELST KR (1975) Quantitive Models for Police Patrol Deployment. PhD Thesis, MIT, Cambridge, MA.

CHOFFRAY J-M (1977) A Methodology for Investigating the Structure of the Industrial Adoption Process and the Differences in Perceptions and Evaluation Criteria Among Potential Decision Participants. PhD Thesis, MIT, Cambridge, MA.

CRETIN S (1975) A Model of the Risk of Death from Myocardial Infarction. PhD Thesis, MIT, Cambridge, MA.

ERIKSEN SE (1977) Policy Oriented Multi-Equation Models of US Domestic Air Passenger Markets. PhD Thesis, MIT, Cambridge, MA.

GABBAY H (1976) A Hierarchical Approach to Production Planning. PhD Thesis, MIT, Cambridge, MA.

GINZBERG MJ (1975) A Process Approach to Management Science Implementation. PhD Thesis, MIT, Cambridge, MA.

GORDON JR (1977) The Congruence Between the Job Orientation and Job Content of Management School Alumni. PhD Thesis, MIT, Cambridge, MA.

HOROWITZ M (1975) A Study of Interdisciplinary Education at MIT: The Concourse Program. PhD Thesis, MIT, Cambridge, MA.

KARMARKAR US (1975) Multilocation Distribution Systems. PhD Thesis, MIT, Cambridge, MA.

KRUMLAND RB (1977) Base Concepts and Mechanisms in Knowledge-Based Model Building for Managers. PhD Thesis, MIT, Cambridge, MA.

MALHOTRA A (1975) Design Criteria for a Knowledge-Based Language System: An Experimental Analysis. PhD Thesis, MIT, Cambridge, MA.

PLOVNIK M (1975) Individual Learning Styles and the Process of Career Choice in Medical Students. PhD Thesis, MIT, Cambridge, MA.

Runge DE (1976) Labor-Market Dynamics: An Analysis of Mobility and Wages. PhD Thesis, MIT, Cambridge, MA.

Sen AJ (1977) Faculty Attitudes Towards Governance. PhD Thesis, MIT, Cambridge, MA.

Tushman ML (1976) Communications in Research and Development Organizations: An Information Processing Approach. PhD Thesis, MIT, Cambridge, MA.

Malik MAK (1975) Optimum Capital Planning of Reliable Systems. PhD Thesis, New York University (Available at Libraries of New York University and Polytechnic Institute of New York, and from University Microfilms, 333 N First Street, Ann Arbor, Michigan 48106, USA).

Rock H (1976) *German: Grundlagen und Verfahren der ordinalen sequentiellen Problemeinbettung.* PhD Thesis, Technische Universitat, Berlin (Available at: Universitatsbibliothek der TU Berlin, Abteilung Dissertationen, Strasse des 17 Juni 135, D 1000 Berlin (West) 12).

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 3. 00. 173 to 177
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0601-0173$02.00/0

Vol. 7
1979

EDITORIAL

Lateral Consistency?

I WAS TRADING reminiscences some time ago with this consultant ex-friend of mine (call him JBJ for short) about our experiences in assessing the capabilities of companies to meet new challenges. We had both been involved, one way or another, as observers and as change agents, in studies of evolving corporate strategies, and our combined experience adds up to a fascinating story of contrasts in the behaviour of management, of action and inaction, of innovation and inordinate conservatism, of courage and timidity, of flashes of insight and obstinate blindness, of success and abject failure.

It is a story of a struggle for survival, when new threats and opportunities constantly appear on the horizon: new products that could supersede an existing range, new materials and methods of manufacture that threaten established expensive facilities, new moves by competitors to capture a larger share of the market, new markets and distribution methods, new industrial groupings and changes in sources of funds. All these, and many more, pose problems for the top managerial echelon of the enterprise. Every new challenge can be seen both as a threat and as an opportunity, and a suitable response needs to be formulated. An aggressive reaction often involves an outlay of resources and greater exposure, with dire financial consequences if the move fails; a defensive reaction is generally safer in the short term, but it could hail the beginning of a steady decline.

What is it, then, that makes management take one course of action rather than another? It is easy, of course, to be clever after the event, to argue that the writing was on the wall for all to see, and that only one particular posture should have been adopted. In the absence of hindsight, it is not only the assessment of whatever evidence is available at the time that needs to be considered, but the organization structure and its procedures, as well as the weight of tradition, obviously have a part to play.

Now, said JBJ, in order to understand what shapes the attitudes of decision-makers and their approach to problem-solving, we have to identify the one pervasive characteristic that prevents them from

having an open mind, from taking a fresh and unbiased look at a problem.

And what is that characteristic? Vertical thinking, boomed JBJ triumphantly. Vertical thinking is the product of our educational system, which strives to inculcate rigid forms of reasoning, with emphasis on the methodology of logical deduction. Perhaps the best example of that, JBJ warmed to his theme, is the study of Euclidean geometry, where you start with certain definitions and a given set of reasonable premises, and you then proceed to prove relationships between various properties. And the method of proof assumes a rigid format, which is symptomatic of this deductive approach: You define 'the givens', you make a statement in the form of a theorem to describe what it is that you seek to prove, and then you use other proven theorems (whose proof does not depend on the one before you) to arrive at the desired conclusion, ending with the punchline QED to put the final stamp of approval on the rigour of the method and the validity of the outcome. Thus, the whole structure of Euclidean geometry consists of bricks that lean on other bricks, a vertical monument of deductive logic, which is at pains to separate premises from arguments and to highlight the principle of economy in the development of your reasoning (I vividly remember the occasion at school when I was called to the blackboard to prove a new theorem and was admonished by the teacher for "going back to first principles", as he put it, instead of relying on another theorem, which had just been proved and which would have saved two lines in the proof I had proposed).

This kind of vertical thinking is deep-rooted. It emanates from conventions about approaches to problem-solving, so that our thought processes are geared to follow systematic analyses, where every exploratory step has its place in a sequential and inter-dependent set of arguments. Even the problems that we need to solve are influenced in their formulation by the vertical thinking process, and once a problem is formulated in a particular way, we are conditioned to seek a solution accordingly, whereas lateral thinking allows us to abandon the rigid formulation and to pursue unconventional avenues. One of the well known elegant examples cited for the contrast between vertical and lateral thinking concerns the case of organizing a tennis tournament with 100 entries, where the winner of each match goes forward to the next round, and the problem is to determine the number of matches that need to be scheduled to produce a champion. The conventional way of approaching this problem is to concentrate on the winners and to work out the number of matches in each round, a somewhat laborious process, whereas if the problem-solver turns his attention to the losers he immediately perceives that the purpose of each match is to produce one loser and the purpose of the whole tournament is to produce 99 losers, hence 99 matches need to be scheduled.

It is examples like this that JBJ kept trotting out to illustrate how backward and old fashioned vertical thinking was. I surmised that he must have attended one of these lateral thinking courses, probably run by that smooth and articulate proponent of lateral thinking, Edward de Bono. And I was right, he had been to one of these courses, and caught the bug. It was clear to him that all our ills would be cured if we were to adapt in industry a positive approach, if we were to brush aside all conventions that are only designed to impede our progress, and allow ourselves the luxury of examining problems with a fresh outlook, without constraints or inhibitions. Constraints, he muttered, there is the epitome of the conventional approach to decision-making, to find a solution subject to constraints. Constraints are but artificial barriers in our exploration, they compel us to concentrate on them, they drive us to thinking in terms of how to circumvent them or how to minimize their effect, whereas what we should be doing is freeing ourselves of such shackles and be prepared to consider any bold solution, however atrociously bizarre, and be equally prepared to abandon at a stroke set procedures and existing plans.

I could not deny the strength of these sentiments. Organizations do tend to become conservative and bureaucratic, so that some iconoclastic approach is necessary if an innovative impetus is to be sustained. But, then, how can we reconcile this exhortation for lateral thinking with the need for consistency in organizations?

Explain, demanded JBJ. Well, the argument goes something like this: Society requires that actions of its members, both of individuals and of groups, be carried out according to some rules. Hence we have a framework of legislation, that prescribes a series of do's and dont's, and people are expected to obey the law. In addition, there are unwritten laws, which equally affect the behaviour of people, so that those who do not conform are regarded as eccentrics. And it is because we want to be able to predict to a large extent how people and organizations will react in given circumstances that we often refrain from putting real eccentrics in charge. Consistency and order are the foundations on which organizational structures strut. Without consistency, you don't know how your subordinates will function, you don't know how your superiors will handle new problems, you don't even know what their expectations are, in short you don't know where you stand.

I have seen this state of affairs in one firm, where the chief executive kept on changing his mind about various major issues which needed attention: Manager *A*, who was supposed to be in charge of all personnel matters, suddenly found that *B* was instructed to deal with recruitment, while *C* was issuing circulars about working conditions; one day *A* was specifically told that a certain executive was to have a personal assistant recruited to help him, but when all the interviews were over and a final short list compiled, a decision was made to

abolish the new post; then, a week later, *A* found that an employee had been transferred from another department to fill the non-existent vacancy, needless to say, with the chief's approval. To overcome these difficulties, it was decided as a matter of policy that job specifications would be compiled for every managerial position, so that lines of demarcation of responsibility and accountability would become clear, but no sooner was a job specification finalized, it became partly obsolete because of a constant redefinition of responsibilities and changes in the organization structure. Who are you reporting to today? was the standard joke in the company. Now, every single decision made by the chief executive had a reason, every change of mind was the outcome of new information or in response to new stimuli, but to an outsider the general impression was that of a dynamic organization steered firmly without its members having a clear sense of direction. The only consistent and predictable feature of that style of management was that there was no consistency, and that every final decision was never final.

I put these points to JBJ and he conceded that there was a problem, and that lateral behaviour was often incompatible with consistent behaviour. But in trying to determine which style of management is more beneficial, a stodgy-bureaucratic approach that upheld consistency at all costs or an eccentric-unpredictable mode of control that kept everybody on his toes, we need to look at the success rating of the companies in question. Was the firm that I alluded to, JBJ demanded, the one with the dominant chief who kept on changing his mind, successful or not? I admitted that it was financially very successful, that the managerial staff was remarkably adaptive in accommodating the stream of instructions emanating from the top, and that as a result the organization was very well equipped to respond rapidly to changing market circumstances, much more so than its larger competitors. Size has undoubtedly something to do with it; small firms can thrive on weak administrations, while large firms have to enshrine formal procedures, which become their very lifelines.

There you are, JBJ was jubilant. This is why innovation is so difficult to implant and to sustain in large organizations. It simply does not fit in, it is too alien to the whole ethos of their management style; it is even threatening, to the structure and to the individuals who feel that they have most to gain from the status quo. And, he continued, this is why I so often come across firms which are totally moribund, stuffed with replicas of organization men who will not or dare not sweep aside the monuments of convention that surround them. This is why, he concluded, I will not give you two figs for your consistency.

It was sometime later that JBI suggested that I join him in a new venture that he was about to launch. Since at that time I was involved with a competitor I had to consider carefully the pros and cons of

JBJ's proposition before severing my existing association. His scheme was sound, his terms were good, and I like the guy. I therefore accepted his proposal and gave notice to the competitor, only to discover a few weeks later that JBJ had had a lateral thinking session with himself. He completely recast the new venture, its aims and scope had changed, so had the array of his associates, and my participation was written out of the new blueprint. He knew full well that he was placing me in a very awkward position, but he felt no sense of commitment or obligation, and that is how JBJ became a good ex-friend of mine. Loyalty is the product of consistency, and it has no place in the scheme of those who despise it.

The dilemma, then, remains, the dilemma of how to achieve coexistence between consistency and innovation. No organization can survive for long on the one without the other. We need both; in fact, we need a structured way in which the unstructured can be tolerated and brought to the fore.

SAMUEL EILON
Chief Editor

Vol. 7
1979

BLANK
PAGE

Vol.
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 3. pp. 179 to 182
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0601-0179$02.00/0

FEEDBACK

Ackoff's Fables

I AM VERY grateful for the amount and quality of attention the Editor has given in his Editorial to my book, *The Art of Problem Solving* (see Omega, Vol 7, No 2). He has perceived my intentions and communicated the spirit of the book much better than I could have.

His invitation for comments on the Editorial is generous. I accept with the hope that he will be able to publish this response in its entirety to avoid misconceptions. As he knows, it is unusual (and it is certainly not my practice) to respond to reviews of one's books, even to unfavorable reviews. I feel reviewers have a right to express themselves without having to defend themselves against offended authors. In the case of such a sympathetic and understanding review as his, it seems inappropriate to dwell on minor disagreements that I may have with it. There are two points, however, on which brief comment may help readers.

First, I refer to his discussion under the heading: 'Cause and Effect.' In the second example, on smoking and cancer, he refers to the enigma facing government and justifies its taking action against smoking by invoking consideration of type I and type II errors. However, he formulates the type I error in terms of "action being taken but having no positive results." This does not take into account the possibility that such action may have *severely negative results*. There is some evidence to this effect with regard to the discontinuation of smoking which some governments choose to ignore or suppress. I recall (but cannot retrieve) a study carried out, I believe, by the Tavistock Institute that showed that those who stop smoking have a much higher incidence of mental illness that those who continue. (This does not establish a causal connection between the two but it gives reason to pause.) Then there are the impressive (to me) studies by Richard J Hickey and Richard C Clelland; for example, 'Maternal smoking, birth weight, infant death, and the self-selection problem,' *American Journal of Obstetrics and Gynecology*, 131 (Aug. 1, 1978), 805–811. Among other things their work reveals the possibility that nicotine induces "a physiological response that serves to alleviate bioenergetic deficiencies in some individuals."

If advertisers in the first of my examples cited in the Editorial were to use the logic Eilon suggests for government in the second example, then, in the absence of causal information, they would advertise more. But this ignores the cost and harm of excessive advertising. Jim Emshoff and I have shown this to be the case where beer is concerned (Advertising research at Anheuser–Busch, Inc. (1963–68), *Sloan Management Review*, Winter 1975, 1–15). Excessive advertising may decrease sales, which is not equivalent to "the lack of positive results." In the absence of causal information it may be very dangerous to assume that an effect is a monotonically increasing function of a hypothesized cause.

Furthermore, and this is really an aside, I suspect the motives of a government such as my own that tries to discourage smoking on the one hand (quite unsuccessfully), and, on the other, succeeds dramatically in subsidizing the growing of tobacco.

The second point, a minor one, involves Eilon's use of my example involving a consumption tax. I completely agree with his conclusion to that discussion: "What is ingenious and exciting to some is commonplace to others." I do not think, however, that my example and his counter examples—our sales tax and your value added tax—make his point. He has isolated the proposed consumption tax from the rest of the system of which it was part. The fact that this tax was imbedded in an Electronic Funds Transfer System eliminated the need for all checks and most cash, and made it possible and desirable to eliminate (1) all income taxes, (2) interest paid by banks on savings, and (3) interest paid to banks for loans, among other things. The system with these characteristics, I believe, is *not* commonplace. If it appears so, then my presentation of it is seriously deficient.

I appreciate his thought-provoking editorial and hope that these brief comments do not make my appreciation seem less than it is.

RL ACKOFF

University of Pennsylvania
414 *Vance Hall*
3733 *Spruce Street*
Philadelphia
PA 19174
USA

A Question of Size

IN MY research work, I have found it useful to make the following conceptual definition.

Economy of scale in research and development including preproduction
A

B	C	D
Economy of scale in plant construction	Economy of scale in plant operations	Economy of scale in marketing

E
Economy of scale in technology

Section A: contribution to product development and production aids, such as tools, molds, programs and design foundation. The evermore complex conditions of many product lines/production areas have increased the smallest volume that can manage a fairly advanced product development within a reasonable amount of time. For already existing production programs there are three important areas of scale.

Section B: plant construction with productive equipment in the form of machines and equipment. An economy of scale clearly defined for most sectors, with a scaling factor of 0.4–0.8. A thoroughly investigated area with both a good technical as well as economic level of penetration.

Section C: plant operations including organization and management of the continuous operations, in which the combined action of equipment, product, economic and human resources is an important element.

Section D: marketing includes distribution and the entire market organization in all markets. These may be wholly-owned subsidiaries, agents, business concerns or other distributors. In my opinion an area difficult to describe, since it has been customary to handle joint distribution and trade differently for different products. There are probably considerable advantages in an economy of scale if a second product has its own worldwide distribution. Joint distribution changes the scene drastically in this area.

Section E: Technology means a lot, especially since the technology in any field goes through great changes. An example of this is the replacement of mechanical control systems by electronic control systems using microcomputer technology. The control equipment may be found in the product, in the production or within the organization of distribution. The technology involved may apply to the manner in which products are developed, manufactured and marketed, the organization of the production or the market organization.

Beckett's paper (see Editorial, vol 7, no. 1) is clearly limited to section C, i.e. economy of plant operations/production facilities operations.

STEN OLOF GUSTAVSSON
Chalmers University of Technology
Division of Management
Fack
S-402 20 Göteborg
Sweden

Vol. 7
1979

Political Man

THE STIMULATING editorial 'Political Man' emphasizes the need for '*éducation permanente*' if today's manager is to develop the social, economic, and political skills that tomorrow's corporation will demand. However, I am concerned that management education should help him integrate *all* these facets, rather than create a taxonomy in which he may feel he has to choose a particular role as economic, social, or political man.

Management education—indeed, all education— in classifying and conceptualizing, often seems to veer towards polarizing along a single scale things that might be more effective on separate scales. A company needs to meet economic criteria and social responsibility demands; we have to cope with technology *and* employment; we need to explore ways to meet national *and* multi-national interests; we need the efficiency of unitary control *and* the motivating benefits of ownership and participation; and managers need to realize that reactive *and* proactive roles are two sides of the same coin.

We have myriad examples in industry of managers with this integrative ability, though they seldom have time to express it verbally. One of IBM's secrets of success is certainly its ability to create among its employees both a high sense of owning their own idea or projects, and an equally high sense of membership in the work group as well as the corporation. Similarly Volvo's president doesn't worry whether he is economic man, social man, or political man when he inspires a factory that eliminates assembly lines; neither his motives, his justifications nor his actions gain anything from the segregation. The 'technology versus employment' contradiction has also been eliminated in instances where employees, guaranteed that they will lose neither job security nor income, are encouraged to consider the long-range manning and market implications of new technology, and help find new directions for growth. (Some Swedish newspapers have made progress here.)

One positive contribution that management education might make to this self-managed learning trend, or '*education permanente*', could be an increased awareness that tools, techniques, or 'solutions' must be adapted to specific national, corporate, working group, and individual personalities or cultures. In an era where we are in danger of swinging pendulum-like along a single scale from 'systems' to 'behaviour' this could be a positive, integrating step away from polarization.

NANCY FOY
Oxford Centre for Management Studies
Kennington
Oxford OX1 5NY

In Defense of Determinism

TWO RECENT OMEGA editorials [1, 2] have seriously questioned the validity of attempts to predict structural or managerial characteristics in organizations. While the editorials specifically criticized the Woodward technology-structure model [9] and the Vroom–Yetton leadership model [10], both articles were generally sceptical about the development of models describing or prescribing relationships within organizations. Having recently published research articles questioning Woodward's model [3] and supporting other determinants of structure [4], I agree with the Editor's view that the two models are inadequate while, at the same time, arriving at somewhat different conclusions. To succinctly summarize their shortcomings, the models fail to include relevant predictors; variables are operationalized inadequately, although the Woodward model is decidedly superior to the Vroom–Yetton model in this regard; explanations of linking mechanisms are nonexistent; the models are applied to broad categories of organizations rather than specific types; and both models are presented prematurely for prescriptive (normative) purposes. Acknowledging these problems, it is too early, on the basis of accumulated research, to accept or reject either effort. Therefore, an explanation of each of these shortcomings follows, with suggestions for future work in the further development of such models.

Woodward's original presentation of technology as the single determinant of structure placed her in the vulnerable position of explaining all variation in organizational structures simply with knowledge of the incorporated technologies. As stated in the editorial, this explanation was "too simple, too static, too pat" [1, p. 499]. A failure, then, to accurately predict the structural characteristics of a number of assorted organizations on the basis of a single technological dimension obviates her model. Such grossly reductionistic stances in the social sciences are easily challenged empirically because the world is multivariate in nature and efforts to explain phenomena in bivariate terms often fall short on the basis of the dearth of information provided in dyadic models. Oversimplified models may be appealing conceptually, but parsimony (in theory) should be achieved only by deleting variables on empirical grounds.

The Vroom–Yetton model includes multiple situational predictors of 'management style,' but excludes *a priori* the personality of managers and institutionalized roles as affecting that style. The assertion that a manager's approach in solving a particular problem is situation specific assumes that human behavior is overly rationalistic and represents as reductionistic a position as the Woodward model. The variation in management style, ranging from participative to authoritarian, is presented as a function of the nature of the problem and the characteristics of the work group. If the situation varies within an organization, or between organizations, as a result of heterogeneous problems and/or varying characteristics of the work group being managed, then the style would be expected to change in a specific direction regardless of the manager's particular personality or role within the organization.

In research designs which possess only assigned factors with intact groups, as is the case with most organizational research, the control of confounding effects (otherwise accomplished experimentally by random assignment) must be achieved statistically by including all significant predictors in the model. The predictors are included in the model on the basis of a stated theoretical understanding of, in this case, the determinants of structure or management style and are deleted only when found to explain no variation in the criterion. There are obviously multiple sources of managerial and structural variation in an organization and, therefore, an examination of any single variable's impact must be accomplished within the context of a multivariate explanation of that relationship including those significant sources of variation.

The shortcomings of the Woodward and Vroom–Yetton models underscore the problems facing theorists identifying the determinants of organizational and managerial characteristics. Such theoretical approaches could share the idea that a criterion, centralization for example, is affected by some intra- or extra-organizational influences but disagree on the nature and hierarchy of those influences. Executive personalities as well as worker characteristics might be included as determinants of centralization in such models, along with organizational size, age, budget, *etc.* This contradicts the 'one best way' to organize notion and challenges the 'rules of thumb' administrative approaches founded on the early rational models of organization, but the burden of proof for these more complex explanations lies in the conceptualization of each hypothesized determinant's effect on structure or management style. Within this conceptualization, the specific dimensions of the criterion and its predictors must be clearly operationalized. Lack of clarity results in incongruent operationalizations between research efforts, such as Hage and Aiken's [5] use of Hall's [6] division of labor scale measuring the extent of structural formalization as a measure of technological routinization in changing a raw material. While two distinct variables may be highly related, a dimension of one cannot be correctly operationalized as a dimension of the other.

An issue closely related to the problem of incongruent operationalizations is the 'unrationalized categorization' [8] of multidimensional variables such as structure or management style. In other words, general conclusions about the relationship between structure and another variable cannot be made on the basis of that variable's relationship with one dimension of structure, such as the hierarchy of authority. Confusion results as easily from the lack of distinction between various dimensions of a given variable as from the lack of distinction between separate variables.

Another major shortcoming of the two models is a failure to explain just how the variables are linked in a cause and effect relationship. The descriptions of these linking mechanisms are essential to viable theory development. If A cannot be linked to B in a theoretically convincing manner, empirical support for a relationship between A and B is artifactual. For example, in a recent study of the effect of 'time' on human service organizations, the number of years an organization had been chartered was hypothesized to decrease centralization and formalization because of the relaxing of control functions as the assurance of organizational survival increased [4]. In contrast, the average number of years workers had been with an organization was hypothesized to increase centralization and formalization as the 'well-worn paths' of problem solving described by March and Simon [7] are repeatedly traversed [4]. Each variable is a dimension of organiza-

tional 'time,' but each is distinctively and uniquely linked to centralization and formalization. The validity of such conceptual links is as important an issue in evaluating theory as is the assessment and analysis of observed covariation between variables.

In addition to the above requirements, models of relationships between organizational variables must be delineated for defined categories of organizations. For example, I have developed a model of the relationship between technology and structure in human service organizations which directly contradicts Woodward's thesis [3]. The model defines worker discretion as an intervening variable which explains the impact of structure on technological routinization. This worker discretion operates within the context of the current state of the art of those human services in the sample and, for that reason, would not play the same role in more highly developed technologies based upon, for example, greater knowledge of the raw material. So, it seems organizations have been defined too broadly and a greater emphasis on the relationships between particular variables within given categories of organizations is needed.

In summary, descriptive models such as those provided by Woodward and Vroom–Yetton can potentially establish the multiple determinants of selected criteria if variables are adequately operationalized and the mechanisms which link the variables are clearly conceptualized for specific types of organizations. If predictors in the models can then be shown to empirically explain a significant portion of variation in the criteria, a contribution to the development of theory is made. Once the validity of a descriptive model is established, prescriptive suggestions for the manipulation of the determinants can be derived on the basis of the desired value for the criterion. In other words, the prescriptive model rests on an understanding of the relationships established in the descriptive model. The presentation of a prescriptive model preceding the establishment of a valid description of the relationships of interest is based on speculation and cannot be viewed as offering serious normative suggestions. While the present deficits in the Woodward and Vroom–Yetton models prevent the realization of either descriptive or prescriptive goals, the guidelines offered here can contribute to the development of similar models which may achieve both objectives.

REFERENCES

1. Eilon S (1977) Structural determinism. *Omega* **5**, 499–504.
2. Eilon S (1978) The Vroom–Yetton leadership model. *Omega* **6**, 469–478.
3. Glisson CA (1978) Dependence of technological routinization on structural variables in human service organizations. *Admin. Sci. Q.* **23**, 383–395.
4. Glisson CA & Martin PY (1979) Productivity and efficiency in human service organizations as related to structure, size and age. *Academy of Management Journal*, in press.
5. Hage J & Aiken M (1969) Routine technology, social structure, and organization goals. *Admin. Sci. Q.* **14**, 366–375.
6. Hall RH (1961) *An empirical study of bureaucratic dimensions and their relation to other organizational characteristics.* Doctoral dissertation, Ohio State University.
7. March J & Simon H (1958) *Organizations.* John Wiley, New York.
8. Stanfield GG (1976) Technology and structure as theoretical categories. *Admin. Sci. Q.* **21**, 489–493.
9. Woodward J (1965) *Industrial Organization—Theory and Practice.* Oxford University Press, London.
10. Yetton PW & Vroom VH (1978) The Vroom–Yetton model of leadership—an overview, in *Managerial Control and Organizational Democracy* (Eds King B, Strenfert S & Fiedler FE), 133–149. John Wiley, New York.

CA Glisson

School of Social Work
University of Hawaii
Hawaii Hall
Honolulu
Hawaii 96822

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 3. pp. 183 to 190
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0601-0183$02.00/0

Handling Financial Risks

PETER G MOORE

London Business School

(*Received September 1978; in revised form December 1978*)

INTRODUCTION

A COMMITTEE to Review the Functioning of Financial Institutions in the United Kingdom (referred to as the Wilson Committee after its chairman, the ex-Prime Minister) was set up in January 1977. A great deal has been written and said in the evidence given to it concerning the handling of *risk* when considering the financing of investment decisions. However, the picture that emerges is somewhat confused and this paper aims to consolidate the situation, and to examine a number of pertinent issues that are raised thereby.

The word risk is used in a variety of ways by different writers:

(a) In common insurance jargon, risk is the *maximum loss* that an insurer can possibly incur when he accepts some proposition. In this sense it has nothing to do with the chance of that loss incurring, only the level of maximum loss that can be incurred.

(b) A second meaning of the word risk is more closely related to chance, for example when talking of an application for a loan to fund some venture. A *risky opportunity* in this sense is one whose *chance of success* is rated low. Gilt-edged stock is referred to as 'risk-free' implying that the probability of the loss of capital or failure to pay interest when due (in a money sense) in such circumstances is taken as zero, in contra-distinction to the loan stocks of industrial companies where both the capital put up by the investor and the interest payments contain an element of risk.

(c) A third use of the word risk is that contained in the phrase *risk premium*. Suppose that a house is worth £20,000 and the chance of its being destroyed is 1/1000 in a given year. Then the true, or pure, premium to cover this risk is £20,000 × 1/1000 or £20. However, on shopping around, the minimum premium quoted by an insurance company is found to be (perhaps) £35. If you accept this quote, the difference (£35–£20) or £15 is sometimes referred to as the *risk premium*. In this sense it is the excess, over the true or pure premium, that you are prepared to pay to avoid the possible loss of £20,000 occurring. (It should be noted that the term risk premium is also used on occasions as being equal to the total quoted office premium itself, i.e. the pure premium plus the risk premium defined above.)

In this paper we use the word *risk* in the sense of para (b) above, namely as synonymous with the chance or probability of failure or loss occurring. The term *loss* then relates to the consequence of the occurrence of a failure. The term *risk capital* is reserved for the situation where mony is advanced in a situation having risk (i.e. probability of loss greater than zero) without security to the lender.

RISK AND RETURN

The cost of money to the borrower (or return expected by the lender) varies according to the class or degree of risk involved. Thus the rate of return required by the lender is lowest on a gilt-edged or risk-free situation. There is then a steady progression upwards in the interest rate required. Local authority loans are regarded as virtually risk-free, but are not state

guaranteed, and hence require a slightly higher rate of interest. (The failure of the Mersey Docks and Harbour Board a few years ago where the Government did not underwrite the capital is an illustration in point). The debentures of first-class industrial companies will come next, followed in turn by those for second line companies. Returns on equity (ordinary shares) will be higher still, but these are complicated first because they are not fixed, and secondly because the return is basically a mixture of capital and interest. But the 'expected' return will be higher than for a debenture from the same company as the latter ranks higher in safety than the equity.

It seems to have been well established by the Wilson Committee [3] that where projects are 'risk-free' there are generally adequate funds available to support them, although this will commonly require the lender to provide security at least equal to the loan capital sought in order to make a situation risk-free to the lender. Since both the funds themselves, as well as the terms and conditions on which they are lent, are likely to be conditioned by whether or not the situation is a gilt-edged one, efforts will naturally be made to ensure that this is seen to be the case whenever possible. Such desires are well illustrated by situations such as those relating to the National Enterprise Board and British Leyland where 'letters of comfort' [9] were written to assist in obtaining commercial funds for the latter on advantageous terms.

Small firms with few or no assets, but good ideas, can often only arrange funds by individual guarantees which convert the loan to a risk-free status, but nevertheless consequentially place a burden on the guarantors. At one level the onus can be put back entirely on to the borrower, e.g. a bank advancing on a situation with risk attached doing so only against security or a guarantee, which effectively removes the risk from the lenders and leaves it with the borrower. This does, of course, presuppose that the realisable value of any assets pledged matches the bank advance. Banks have suggested that this is not always the case when it comes to the crunch and, to this extent, the bank is taking some of the risk from the borrower. It seems to be a further common maxim that a bank (or other lender) will advance half the money required (on security) if the borrower provides the other half. The lender regards such a procedure as one that guarantees complete motivational commitment on the part of the borrower in both risk-free and risk situations. A prudent man (or lender) spreads his risks, but it would seem illogical to expect a man who has his know how, but no spare capital beyond perhaps that sunk in his house, to risk all on one venture. Such an approach towards lending means that it is commonly easier, cheaper and safer to take over an existing small company than to start up completely from scratch, since it is then simpler to provide the guarantees required in the form of bricks and mortar as well as to demonstrate a track record for the organisation.

THE TIME FACTOR

Risk statements must have their units made explicit. Thus the statement that 'at 1 in 1000 a risk becomes unacceptable' [1], can be meaningless unless we are clear as to whether it is a unique once and for all happening, or it is something that could occur at any moment over a period of time. For example if we say that the risk of a motor car being involved in an accident is 1 in 1000 do we mean per quarter. per 100 hours of exposure on the road, per thousand miles travelled, or what?

The same is true of financial risks. If a project is assessed as having a 1 in 100 chance of complete failure with the loss of the £100,000 capital involved the time scale is important. Suppose the whole project is to last one month and the risk free cost of the loan is 12% per annum or 1% per month. The cost of the loan might then be an administrative charge A plus 1% of £100,000 (or £1000) plus a risk premium of a further 1% of £1000 giving a total cost for the month of A + £2000. Suppose alternatively that the project period is one year, with the same assessed risk of failure occurring in the whole period. The true cost now would be

$$A + £12{,}000 + £1000 \text{ or } A + £13{,}000$$

and not

$$A + 12 \times £2000 \text{ or } A + £24{,}000.$$

Conversion of variable loss amounts and levels of risk into one charge to be added to the interest rate is a normal pre-occupation for an actuary, but is commonly dealt with in commercial

Vol. 7
1979

lending by subjective adjustments to the interest rate required rather than by considering the risk and time periods separately.

A refinement of this situation occurs when the 'amount at risk' and the 'level of risk' both vary over time. The maximum possible loss may only occur over a relatively short period of time whilst in other periods the possible loss is considerably less. Neglect of time consideration can lead to over-emphasis on the maximum risk situation and the cost to be incurred by carrying such a risk overemphasised.

RISK INVESTMENT

Once an element of risk is present in a proposal, ways of handling it must be sought. First a distinction should be made between the various types of finance required and the amounts of each that are at risk. The three main types are as follows:

(i) money to pay for capital equipment;

(ii) money to cover operational losses in start-up period—e.g. salaries uncovered by sales;

(iii) money to fund receivables (i.e. working capital).

The relative proportions of these three kinds of money varies enormously from one situation to another. Setting up an hotel might have a high proportion under (i), with much smaller proportions under (ii) and (iii). Setting up a small consultancy service would have high proportions under both (ii) and (iii) but virtually nothing under (i). If such ventures fail, type (i) money would be lost in part, type (ii) completely and type (iii) perhaps only in very small part. Should the various items be lumped together indiscriminately the total funds advanced are equated to the total amount at risk and the latter thus appears to be rather higher than would be the true situation. Figure 1 shows the profile for a start-up situation for a new small company formed recently to market a financial advisory service. It demonstrates a situation where the amount outstanding at risk is always less than the total amount of funds required (due to creditworthy debtors outstanding), whilst the maximum loss level only exists for a relatively short period.

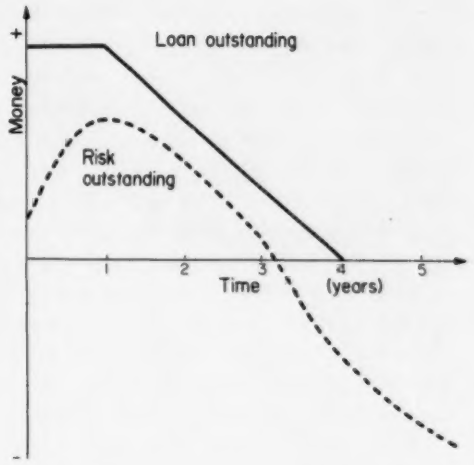


FIG. 1. *Money on loan and at risk.*

UK institutions sometimes appear to treat the various types of money as if they are the same, whereas US institutions are rather more specific in separating out the different categories. It is true that there are specialist services (factoring) available to cover item (iii) but these are sometimes difficult to mobilise for the very small firm dealing with a multiplicity of clients.

A second difficulty occurs when the ratio of (ii) to (i) is high, e.g. in setting up a small consultancy type of service, so that there are virtually no assets to pledge as security for the loan. This is a severe brake on the development of service type organisations as opposed to manufacturing organisations. It is an area that remains very much open for further action, particularly having regard to the view commonly advanced that growth in the economy is likely to rest primarily with the service sector.

THE DIMINUTION OF RISK

There is a natural desire, on the part of both borrowers and lenders, to eliminate or at any rate reduce, risk in any financial venture. Efforts to achieve this can proceed in three main ways.

(1) First, the project can be subjected to detailed and extended preliminary study e.g. by building prototype models, by market research activities or even pilot oper-

ations. Such measures could reduce risk but usually not completely eliminate it.

(2) A second way in which risk can be reduced is via a multi-stepwise decision system. For example in exploration work for minerals one large firm put the initial odds against a profitable mine at 5000 to 1 when a concession on land is taken up after a rough geological, economic and political scan. If geochemical work then done is promising the odds improve to 1000 to 1 against. The next stage is preliminary drilling and, if this is successful, the odds improve further to 25 to 1 against. Finally, if the project stage of detailed drilling is successful the odds narrow to 4 to 1 (2). This is again a situation in which the risk is reduced, but not completely eliminated.

(3) A third way is to look round for a means of sharing or pooling the risk with some other organisation or individual. This can be done in a number of ways, e.g. partial government aid, or a special issue of shares, or some form of joint venture undertaken by a consortium rather than an individual company. The aim in such approaches would be to try to ensure that, should the project go sour, the sponsoring organisation is not crippled as much as it would otherwise have been, although normally this can only be achieved at a price that creams off some of the potential gain. Notice, however, that the risk has not been eliminated as such but merely transferred to a greater number of stakeholders. An example of this can be seen in the workings of the Policyholders' Protection Act of 1975. This requires life assurance companies to contribute up to 1% of their premium income in any one year to guarantee payments to the policyholders of any insurance company which fails. Thus the risks are being transferred from the policyholders of the bad insurance company to those of the 'good' insurance companies. Some of the credit guarantee systems used in various countries, such as the USA, for the development of small businesses take the form of a precept on the successful for the support of the failures. A further, and more positive, example of consortium risk sharing is the formation in Scotland of Seaforth Maritime which has effectively been a joint activity which has resulted in participation in a high-risk venture (North Sea Oil) on attractive terms with proven management (4).

LENDING RATE RELATIONSHIP WITH RISK

There is very little hard evidence about the way in which lending organisations relate the cost of lending to their assessment of the risks concerned. We know from an examination of the annual accounts of American banks that their loss ratios (expressed as a percentage of the amount of money lost in a year through default in relation to the average loans outstanding in that year) are around half a percent per annum for commercial and industrial loans, and a little higher for personal loans. These figures vary slightly from year to year and from bank to bank, but the figures quoted appear to be representative. We do not have equivalent figures from British clearing banks, but we know (from discussion) that over a period of nearly 30 years the Industrial and Commercial Finance Corporation (now an arm of Finance for Industry) have found that an average annual provision of 0.6% in their accounts has covered the losses made. Furthermore, informal discussion with a number of international banks in London suggests that it is common policy to make an annual 1% provision against debts occurring in their books, and that this provision has, generally speaking, been on the safe side.

If figures such as these are representative of the UK banking scene, this would seem to imply that banks have an average probability of total loss (or risk) of 0.01 on each individual advance on an annualised basis, although presumably this is made up of a number of situations with varying risks of loss such that the average annualised risk weighted by the size of the loss, works out at 0.01. Furthermore, the interest rate charged would presumably be the rate appropriate to a risk free investment (gilt-edged), plus 1% for risk, plus an administration charge. The Industrial and Commercial Finance Corporation has confirmed in discussion that, once a proposal is accepted, the terms are virtually the same and do not vary

Vol. 7
1979

greatly with any perceived variations in risk amongst the accepted proposals.

Many margins are, it is true, higher than the 1% mentioned above and this is particularly the case with lending to persons where they can rise as high as 12%. Banks have suggested, however, that a more normal range for the risk margin is from $\frac{5}{8}$% (which would be regarded as rather on the low side) to about 4%, the latter being for small companies and persons. The margins are more linked to the category of borrower than the particular project concerned. It would seem that the high interest rates at present occurring tend to leave little room for risk-taking of an equity type which is really more suitable for many high risk ventures (dividends can only be paid from profits generated so extremely high rates of profit are necessary to make it a superior form of investment for an investor). Correspondence with Mr A Gibb of the Durham University Business School suggests that in the clearing bank market there is little scope for taking higher risks compensated for, say, by higher interest rate; the emphasis is on a low risk profile. This may be partly due to the fact that, unlike equity investments, the possible return from a loan arrangement is truncated at the upper end. It is not uncommon, however, for other lenders of loan capital to demand a share of the equity as part of the price for making a loan. A recent small business set-up was achieved with a £10,000 loan from the founders, a £7,500 loan at base rate plus 3% from a company that wanted to be associated who in turn also demanded 41% of the nominal £100 equity. Banks were not prepared to lend without further personal security as the company was in a service industry and would have few physical assets.

Many would be borrowers are not willing to give up large slices of equity as a condition for a loan and do not proceed with their projects for this reason. It has been suggested that loan guarantee systems could overcome such problems, but this can still saddle an organisation with debt and interest payments when it really needs a substantial equity injection.

JUDGEMENTAL GATES

There seems to be some evidence that the market does not operate in a way in which it assesses risks and shades interest margins to correspond, but rather sets up some technical barrier and does not allow items assessed above that limit to receive funds at all.

To illustrate how such an approach might operate imagine a large set of projects (or proposals) being assessed and rated as to expected return by some analytical cum judgemental approach in advance, and that one then subsequently knew what happened, assuming the funds to have been granted. The relationship between the assessed and actual returns would presumably be something along the lines demonstrated in Fig. 2. The vertical axis shows the return achieved if the project concerned went ahead. The frequency distribution of returns has been laid on its side against that axis. Similarly the horizontal axis illustrates the frequency distribution of the prior judged rate of return that would be achieved for each one of the same set of projects. The body of the diagram illustrates the inter-relationship.

Now, ideally the relationship between rating and outcome would be a perfect one falling, say, along the line *AB*. This would enable a rating level to be selected that would guarantee that all projects over this rating would reach some corresponding actual return. (Or alternatively allow the terms of lending to be directly related to the anticipated return.) In practice, the line *AB* might represent only the average rate of return achieved by all the projects rated at some level L, and there would be a distinct spread of actual returns about the line. This spread is indicated by the shading, the thicker shading representing the more common combinations of rating and outcome, the lighter shadings the less common combinations.

If C_0 represents the acceptable level of return that it is desired to achieve, then the judgemental criterion L can be selected to try to meet that requirement. The further the cut off level, L_0, below which projects are rejected, is moved to the right, the greater will be the chance that the achieved return exceeds C_0 and *vice versa*. On the other hand the further L_0 moves to the right the fewer the projects allowed through the sieve and, as a consequence, the larger the number of projects that would have exceeded the C_0 level that are turned down.

An immediate question to ask, therefore, is whether the direct relationship between L and C cannot be improved through better forecasting techniques. This would be equivalent to

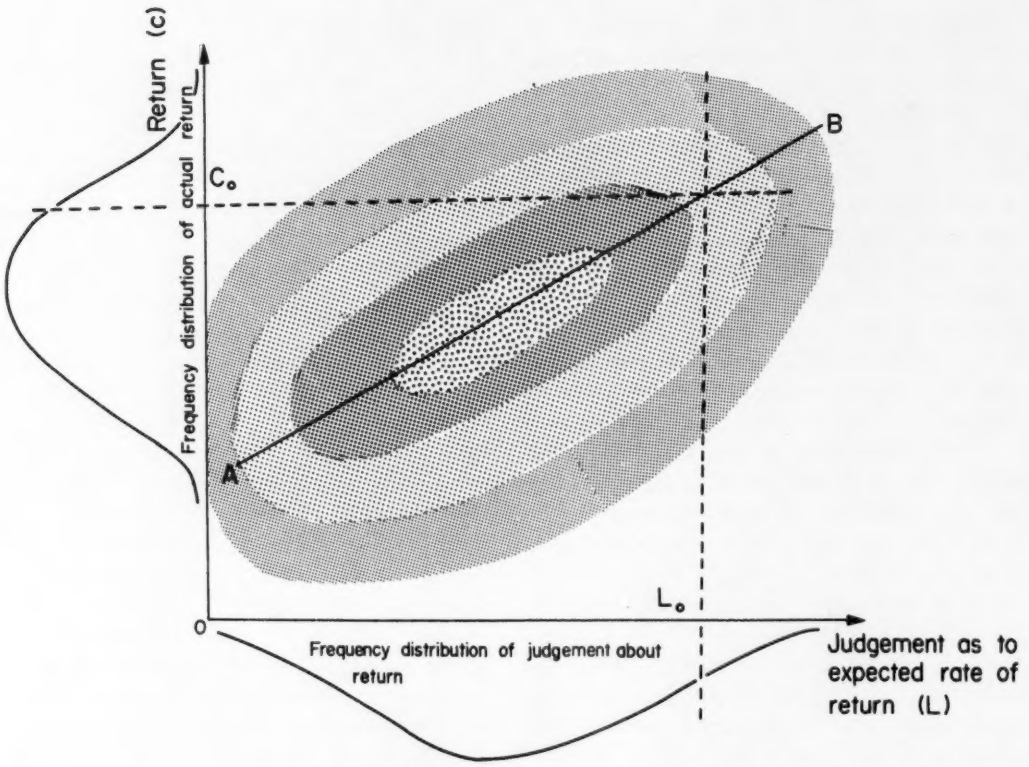


FIG. 2. *Relationship between judgement and return.*

reducing the lesser shaded areas of Fig. 2, and trying to concentrate the shading closely around the line *AB*. If this could be done, then it would be possible to choose, in the ultimate ideal situation, a cut-off level L_0 such that all investment projects that would fall below C_0 are rejected.

Evidence given to the Wilson Committee from a variety of sources suggests that it is general lending policy to keep the risks as low as possible, i.e. that the gate being set is a high one. Whilst the argument for this is primarily that for other cases the required interest rate would be too high to be acceptable to the borrower, it is more likely that the reason lies in the difficulty that lenders have in discriminating between a 1% risk situation and a 3% risk situation (say). Thus the aim of the lender is to train himself first to recognise a risk situation and then to recognise an acceptably low one; he does not feel able to discriminate between various shades of risks. Einham and Hogarth [5] discuss how judgement can be reinforcing—a waiter gives bad service and an inconvenient table to a young couple whom he thinks will tip poorly; when they do so it reinforces his view that young people tip poorly. The same can be true in the judgemental field of assessing risks and allotting resources. Moore [7] discusses some of the difficulties that arise in assessing low levels of risk and methods used to overcome the difficulties.

RISK AND UNCERTAINTY

Industrial practice is to distinguish between the meanings given to the two words *risk* and *uncertainty*. The former term is reserved for situations where quantification of the risk concerned would have a general measure of agreement between a number of experts working in the relevant field. It is then felt reasonable to apply the usual rules of probability to the assessments made.

Common practice is then to separate out other areas of doubt where it is felt that such probability estimates cannot be made on any reasonable and agreed basis; the situations are then labelled with the term 'uncertainty'.

Vol. 7
1979

Examples cited in this context are political uncertainties such as nationalisation, or economic uncertainties such as inflation. Statisticians would argue that there is no real distinction between risk and uncertainty as defined here, and that a distinction is really being made between repeatable and non-repeatable events. Even this distinction is untenable in practice (see, for example, [6] Chapter 2) and hence the notion of two kinds of events is basically unreal. Nevertheless, the distinction persists in the minds of many practitioners and seems to reflect, not so much a challenge to the existence of a probability, but a recognition that in some cases the assessments of probability of some event made by a number of individuals would vary greatly (uncertainty) whereas in other situations (risk) they would tend to have broad agreement.

Inability to reach such agreement does not mean that decisions involving such uncertainties can be eliminated. There is general agreement that uncertainties of the environmental kind affecting business operations have increased enormously. Large firms are particularly affected by a number of political uncertainties; all firms are affected by uncertainties concerning growth, inflation, energy crises, profitability, industrial relations *etc.* Such uncertainties seem to show very few signs of diminishing, and one of the consequences must be that firms will tend to delay investment decisions until they feel that the uncertainties have been reduced to an acceptable (or indeed assessable) level. The asymmetry of returns from new investment makes this caution virtually inevitable, but it has the consequence that there are delays in taking up opportunities that would otherwise have been taken earlier and is one effect of the persistence of inflation at the present time.

Some of these uncertainties cannot be eliminated either by governments or the firms themselves; for example, the market for some product in Africa in five years' time. There are, however, other uncertainties that are more within domestic control, e.g. taxation policies, employment acts, interest rates *etc.* It would seem likely that the more the uncertainty about the latter can be reduced or eliminated the more the rate of investment in industry and other bodies is likely to stabilise at a high level. (See, for example, Nickell [8].) Investment may well be put off to try to time the interest rate minimum correctly as this is commonly fixed for a long time ahead, and tends also to be related to the price at which equity can be raised. The consequence is that peaks and troughs in the economic cycle are exaggerated rather than reduced by frequent Government actions that increase uncertainty.

CONCLUSIONS

This short paper raises a number of issues that are of some concern to those responsible for borrowing or lending money in situations that are not 'risk free', of which the following are the most noteworthy:

(a) Risks cannot be eliminated completely, particularly in a growth economy, but they can be reduced or shared.

(b) The cut-off point for lending may be put at too low a level of risk with consequential diminution of the amount of successful innovation undertaken.

(c) Systems of, and costs of, lending are not generally geared very markedly with the *level* of risk concerned and confusion may occur between levels of risk, amounts at risk and the variations of these over time.

(d) Lenders may require levels of equity participation to be associated with loans that are unacceptable to the borrowers for motivational reasons.

(e) Lending for services (where idea plus work in progress are the main 'security') is not so easy to come by as for manufacturing purposes.

(f) More efforts need to be made to assess levels of risk rather than to concentrate on a no risk/acceptable risk form of classification.

REFERENCES

1. Ashby Lord (1977) The subjective side of assessing risk. *New Scientist* **74**, 398–400.
2. Cheesewright P (1978) Mineral exploration: pay now and hope to collect late. *Fin. Times*, Feb 7.

3. Committee to Review the Functioning of Financial Institutions (1977) Progress report on the financing of industry and trade, HMSO.
4. Committee to Review the Functioning of Financial Institutions (1978) *Financing of North Sea oil.* Research Report No 2, HMSO.
5. EINHAM HJ & HOGARTH RM (1978) Confidence in judgement: persistence in the illusion of validity. *Psychol. Rev.* **85**(5), 395–416.
6. LINDLEY DV (1971) *Making Decisions.* John Wiley, London.
7. MOORE PG (1977) The manager's struggles with uncertainty. *Jl R. Statist. Soc.* **140**, 129–165.
8. NICKELL SJ (1977) The influence of uncertainty on investment: is it important? *Econ. Jl* **87**, 47–70.
9. STEVENSON H (1978) Letters of comfort. *The Times*, Jan. 30.

ADDRESS FOR CORRESPONDENCE: *Professor PG Moore, London Business School, Sussex Place, Regent's Park, London NW1 4SA.*

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 3. pp. 191 to 205
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0601-0191$02.00/0

A Risk Analysis of an LNG Terminal

RALPH L KEENEY

RAM B KULKARNI

KESHAVAN NAIR

Woodward–Clyde Consultants, California, USA

(*Received in January* 1979)

Vol. 7
1979

I. INTRODUCTION

EVEN THOUGH the domestic supply of natural gas is declining, there are strong arguments in favor of its continued use: it burns cleanly; an efficient distribution system exists; and consumers prefer gas as the fuel for heating homes and other buildings. One solution to the declining domestic supply problem is to import natural gas—the liquefication and transportation of natural gas is a well-developed technology. The major concerns about the large-scale importation of liquefied natural gas (LNG) are related to safety. This paper presents an analysis of the public risks associated with operating a proposed facility to import liquefied natural gas.

The El Paso LNG Company and its subsidiaries plan to build and operate a marine terminal and LNG vaporization facility, named La Salle Terminal, in Matagorda Bay, Texas. Matagorda Bay, located approximately 120 miles southwest of Houston on the Texas Gulf Coast, is not a densely populated area. Although there are no cities or major towns in the vicinity of the proposed terminal, there are some small towns (see Fig. 1).

The La Salle Terminal is designed to process one billion cubic feet of natural gas per day. This LNG, which will be obtained from Algeria, will be delivered to the terminal by a fleet of LNG carriers. There will be approximately 143 LNG carrier arrivals in the facility per year, and each carrier will deliver approximately 125,000 cubic meters of LNG per arrival. After an LNG carrier berths, the LNG will be transferred via cryogenic piping to three LNG storage tanks, each of 100,000 cubic meters capacity. The LNG will be independently withdrawn from these tanks, revaporized, and then piped to consumers via a high-pressure intrastate pipeline.

All phases of the design, construction, and operation of the La Salle Terminal and LNG Carrier Fleet will employ special safety procedures and techniques. The La Salle Terminal LNG storage tanks will be located in diked areas and will be designed to minimize any possible spillage. The LNG carriers are of special double-hull and double-bottom construction, and will use sophisticated anti-collision and navigational systems. Special US Coast Guard operating procedures which apply to LNG carriers in US waters and Matagorda Bay will be followed. Fire-protection and control equipment designed to deal with LNG fires will be constructed and maintained at the facility.

All of these safety considerations are designed to reduce to an extremely low level the likelihood of an LNG accident that could result in consequences to the public. However, such an event is possible. The estimation and evaluation of the risks to the public from the operation of such a facility is an important element in determining the feasibility of the project and in obtaining approval from the appropriate regulatory agencies. This analysis examines the public risks.

Problem statement

The problem statement for the risk analysis of the proposed La Salle Terminal was to

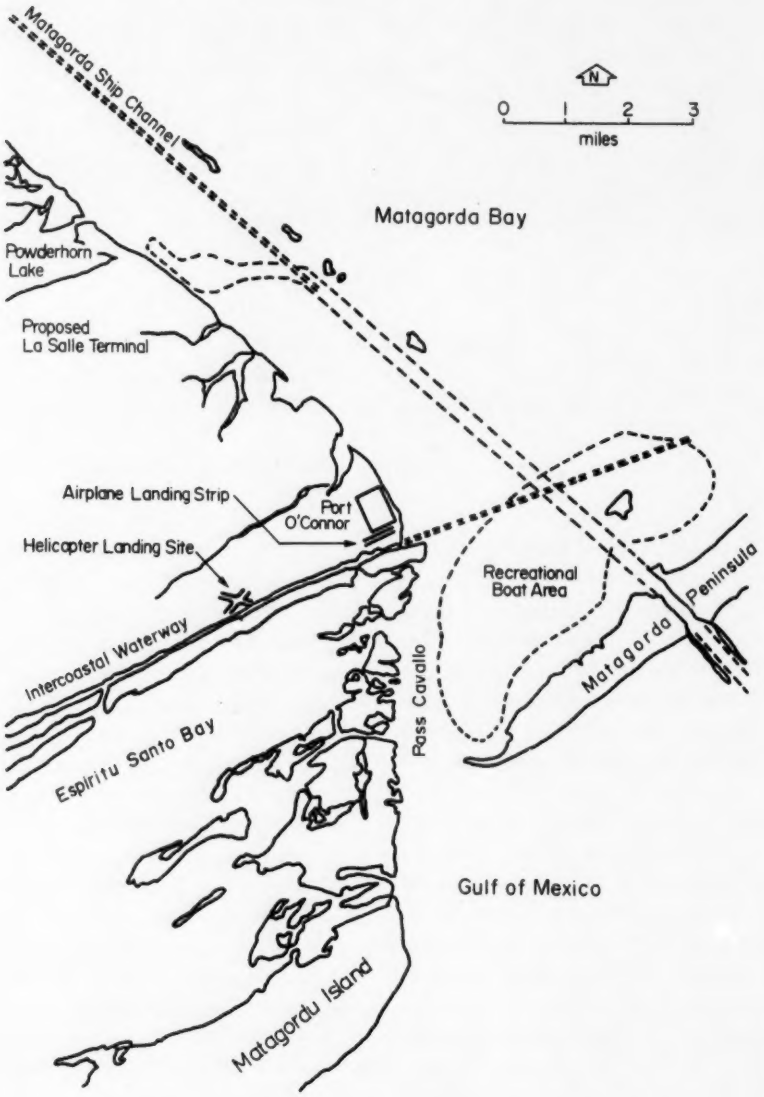


FIG. 1. *The location.*

examine the public risks from releases of LNG due to natural and external man-made events and accidents due to project operations.

To create a potential for public fatalities as a result of the construction and operation of the La Salle Terminal Facility, a series of events must occur in a particular sequence. First, an initiating event such as a ship collision or a terminal-equipment malfunction resulting in a spill of LNG must take place. Because of the nature of the circumstances required to cause such a spill, it is very likely that the spill would be immediately ignited. If it is ignited, a spill pool fire will result. In the unlikely event that the spill does not immediately ignite, the very cold (−260°F) LNG rapidly evaporates, forming a negatively buoyant LNG plume. Under certain weather conditions (defined by wind speed and direction and stability class), this plume may be carried to an area occupied by the public. The mixture of vaporized LNG and air is such that the vapor cloud may ignite as a result of contacting an ignition source, such as a gas pilot light, the flame of a cigarette lighter, or an electric spark. If the vapor cloud does not ignite within a certain length of time it will disperse, as the ratio of vaporized LNG to air will be

Vol. 7
1979

too low to allow ignition. Individuals caught in a spill pool or vapor cloud fire, or exposed to high thermal radiation levels causing severe burns, may become fatalities.

II. APPROACH: THE RISK ANALYSIS MODEL

A schematic representation of the La Salle risk analysis model is given in Fig. 2. The model can be divided into three components:

- Development of accident scenarios and their associated probabilities.
- Quantification of public risks.
- Evaluation of public risks.

Because of the complexities of any risk analysis, simplifying assumptions must be made. The spirit of this analysis required that all assumptions be stated explicitly and conservatively. That is, these assumptions will cause the analysis to over-estimate the public risks.

Development of accident scenarios and their associated probabilities

An accident scenario is a sequence of events that must occur for public risk to exist. Elements 1 through 5 of Fig. 2 constitute development of the information necessary for formulating the accident scenarios. Figure 3 provides a pictorial representation of such a sequence of events.

An accident scenario corresponds to a specific sequence of branches in the event tree of Fig. 3. A representative accident scenario could be described as follows: An LNG carrier collision occurs in the harbor, releasing an LNG spill of a certain size. There is no immediate ignition, so a vapor cloud forms. The wind is from the east at 10 miles per hour with Pasquill stability[1] class *D*. The eighth ignition source ignites the vapor cloud.

As indicated by this chain of unlikely events, each element of the accident scenario must take place either simultaneously or in a specific order. In the risk analysis model, the annual probability of a particular scenario is the product of the following factors:

- Annual probability of the initiating accident,
- Probability of no immediate ignition for that accident,
- Probability of the wind direction,
- Probability of the wind speed and stability class given that wind direction, and
- Probability that the n^{th} ignition source ignites the vapor cloud.

Similarly, the probability of a particular accident scenario which results in a pool fire is equal to the annual probability of the initiating accident multiplied by the probability of immediate ignition.

Accident scenarios were constructed for all combinations of the individual events, and their probabilities calculated.

Quantification of public risks

The purpose of the risk quantification study is to compute the public risks due to LNG terminal operations. Four categories of public risk were examined in this analysis:

- *Societal risk*—total expected fatalities per year.
- *Individual risk*—probability of an individual exposed becoming a fatality per year.
- *Group risk*—probability of an individual in a specific group exposed becoming a fatality per year.
- *Risk of multiple fatalities*—probability of exceeding specific numbers of fatalities per year.

Once the probabilities of all accident scenarios are determined, the probability $\Pr(x)$ of x fatalities is calculated from

$$\Pr(x) = \sum_i \Pr(x/S_i) \Pr(S_i), \quad (1)$$

where $\Pr(x/S_i)$ is the probability of exactly x fatalities resulting from accident scenario S_i

[1]The Pasquill stability class is an indicator of atmospheric mixing potential. The mixing potential decreases from Class *A* through Class *G*. In this study, conditions are such that only classes *A* through *F* are necessary.

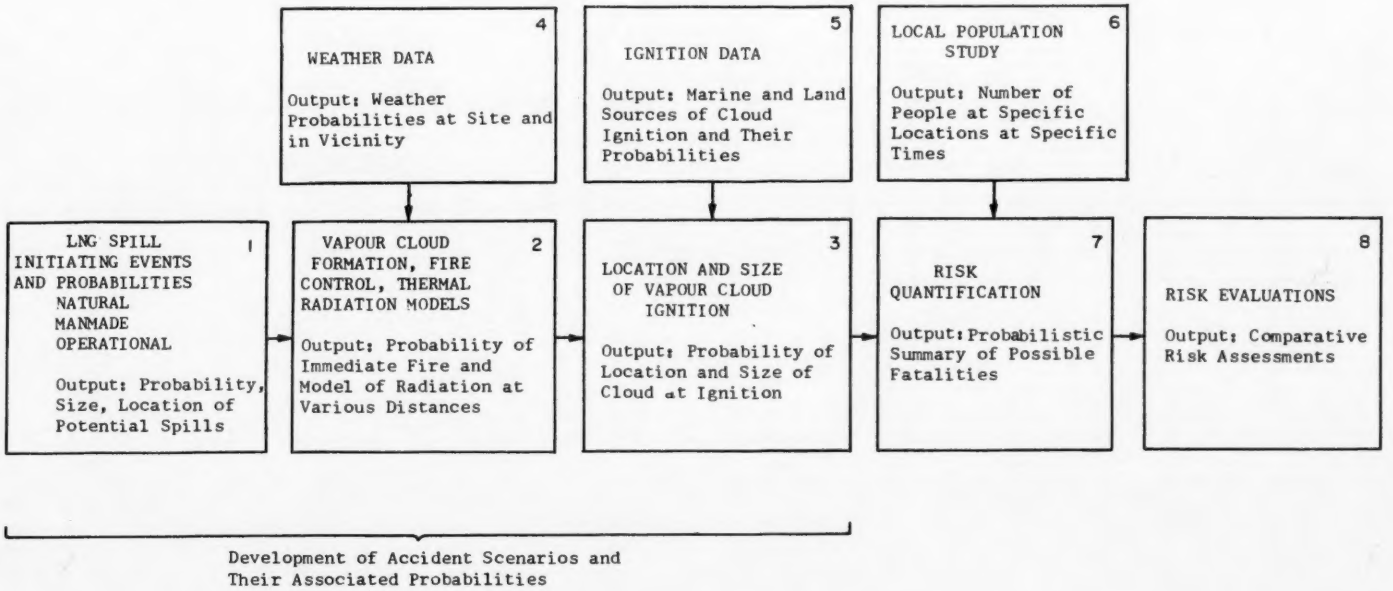
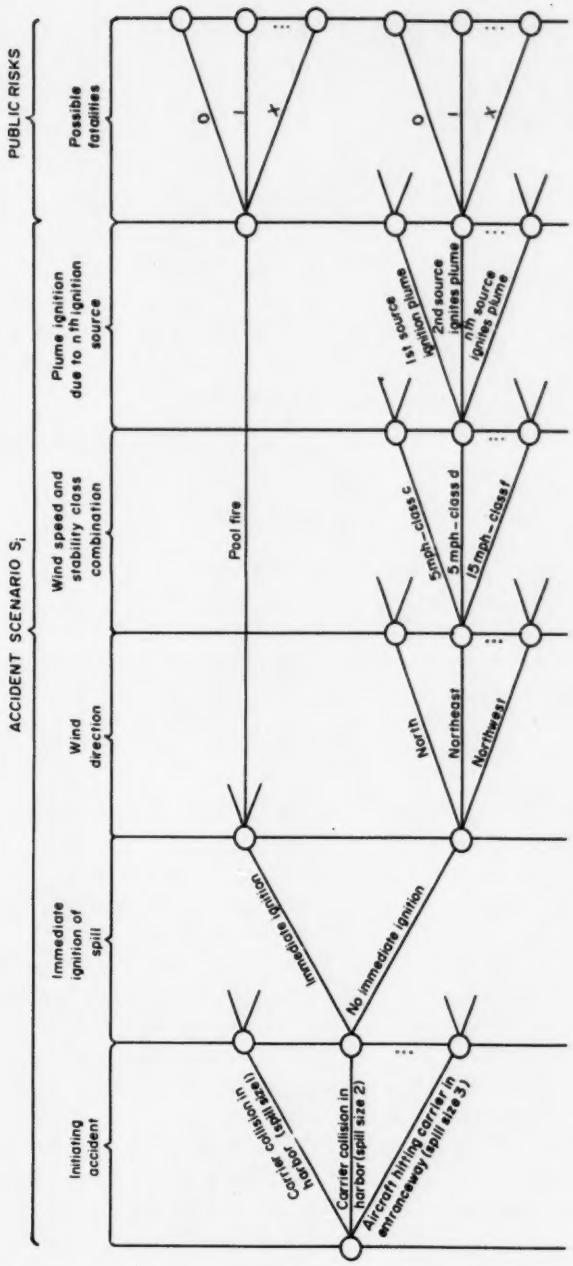

4
WEATHER DATA
Output: Weather Probabilities at Site and in Vicinity
5
IGNITION DATA
Output: Marine and Land Sources of Cloud Ignition and Their Probabilities
LOCAL POPULATION STUDY
6
Output: Number of People at Specific Locations at Specific Times
LNG SPILL INITIATING EVENTS AND PROBABILITIES
1
NATURAL
MANMADE
OPERATIONAL
Output: Probability, Size, Location of Potential Spills
VAPOUR CLOUD FORMATION, FIRE CONTROL, THERMAL RADIATION MODELS
2
Output: Probability of Immediate Fire and Model of Radiation at Various Distances
LOCATION AND SIZE OF VAPOUR CLOUD IGNITION
3
Output: Probability of Location and Size of Cloud at Ignition
RISK QUANTIFICATION
7
Output: Probabilistic Summary of Possible Fatalities
8
RISK EVALUATIONS
Output: Comparative Risk Assessments
Development of Accident Scenarios and Their Associated Probabilities

FIG. 2. *La Salle terminal facility risk analysis.*



Note: At each node probabilities are assigned to the various events

FIG. 3. *Simplified event tree showing events leading to fatalities.*

and $P(S_i)$ is the annual probability of scenario S_i. The expected number of fatalities $F(S_i)$ due to accident scenario S_i is calculated from

$$F(S_i) = \sum_x x \Pr(x/S_i). \quad (2)$$

These formulas provide the basis for quantifying all of the public risks.

Societal risk is indicated by the expected number of fatalities F per year. This is calculated from

$$F = \sum_i \Pr(S_i)F(S_i), \quad (3)$$

where the contributions of all possible accident scenarios are summed together.

Individual risk is measured by the annual risk level R to an individual in the population exposed to possible risks from LNG. This is found by dividing the expected number of annual fatalities by the total number N of people exposed, yielding

$$R = \frac{F}{N}. \quad (4)$$

This risk level is the probability that an exposed individual will be a fatality in a specific year.

If for each accident scenario S_i the expected fatalities $F_G(S_i)$ to individuals in group G is tabulated, the overall expected fatalities F_G per year to group G is found to be

$$F_G = \sum_i \Pr(S_i)F_G(S_i). \quad (5)$$

The *individual risk R_G in a particular group G* is

$$R_G = \frac{F_G}{N_G}, \quad (6)$$

where N_G is the number of people in the group G.

The risk of multiple fatalities is given by the probability that the number of fatalities in a given year is equal to or greater than a specific level y. This can be calculated directly from (1). It is simply the sum of the probability of y fatalities, $y + 1$ fatalities, and so on. Hence

$$\Pr(\text{fatalities} \geq y) = \sum_{x \geq y} \Pr(x). \quad (7)$$

Risk evaluation

There are no generally accepted criteria for evaluating public risk. The approach used in this analysis was to compare risks generated from this project with existing risks to the public and risks from alternate energy sources, and with acceptable levels of public risks suggested in the literature.

III. DEVELOPMENT OF ACCIDENT SCENARIOS AND THEIR ASSOCIATED PROBABILITIES

The elements of this step of the risk analysis consist of determining: (1) probability, size, and location of potential spills; (2) probability of vapor cloud formation and downwind travel and probability of immediate fire; and (3) probability of the location and size of cloud at ignition.

Probability, size, and location of potential spills

The events that can cause spills can be categorized as follows:

- Natural hazards (e.g. hurricanes, earthquakes) which affect the facilities.
- External manmade hazards (aircraft crashes) which affect the facilities.
- LNG Carrier Fleet accidents.
- La Salle Terminal accidents.

Several analyses were conducted to determine the probabilities that each of these events would result in LNG spills. Particular attention was focused on those spills that would have a potential of resulting in public fatalities.

Natural hazards. The potential risks due to various natural hazards, including earthquakes, severe winds, storm waves and tsunamis, and meteorites were investigated. Essentially, none of these natural events contributes significantly to the public risk in comparison to the risks associated with other types of accidents. A summary of the risks from natural hazards follows.[2]

Using conservative estimates, the likelihood that an earthquake would produce a ground acceleration at the La Salle Terminal site large enough to exceed the design specification of the storage tanks is approximately 10^{-11} per year. Even if the tank did rupture with such

[2] Details of these analyses are provided in [1], Appendix A.

Vol. 7
1979

an earthquake, the analysis for the onshore facilities indicates that public risk is essentially nil. There is a higher probability of pipe breaks than of storage tank failure due to ground motions, but the analysis indicates that the results are inconsequential to public risk.

The main source of severe winds in the vicinity of Matagorda Bay is hurricanes. The primary concern with wind is its effect on the storage tanks, since all LNG carriers will leave and remain outside of Matagorda Bay if any winds greater than 60 *mph* are forecast or observed. The tanks are designed to withstand an instantaneous gust of 217 *mph* and a one-minute wind of 166 *mph*. Meteorological data [2] indicate that a one-minute wind of 166 *mph* occurs once every 100 years, which is twice as frequent as an instantaneous gust of 217 *mph*. Hence, it is assumed that a storage tank may begin to fail with probability 1.5×10^{-2} in any given year. However, even if the tank failed completely, either a pool fire would burn on the site or the wind turbulence would disperse the vapor plume before it passed terminal boundaries.

There are no known faults capable of generating significant tsunamis in Matagorda Bay. Extrapolating data from the storm-wave analysis, the likelihood of damage that could result in public risk from terminal accidents was found negligible. Operating policy requires that in storm conditions all LNG ships will leave Matagorda Bay. Since sufficient warning of impending large waves would be available, the possibility that large waves would cause ship accidents and contribute to public risk is assumed to be negligible.

Using data on meteorite-impact likelihoods, the probabilities of a meteorite penetrating a ship's tanks in the entranceway to the harbor or at the pier were calculated to be 3.23×10^{-10}, 5.49×10^{-10}, and 9.32×10^{-9} per year respectively, or essentially negligible, since the probability of the event is over two orders of magnitude smaller than the probability of ship collisions. Hence, further analysis of this possibility was not conducted. The likelihoods of penetration of terminal storage tanks and rupture of terminal pipelines were also calculated. Since potential risks following accidents in these cases would be contained within the terminal boundary, they were not investigated further.

External manmade hazards. The potential risk from LNG accidents initiated by aircraft accidents was analyzed. The likelihood of an airplane crashing into the storage tanks, the LNG pipelines at the terminal, a tanker at the marine-terminal pier, or a tanker in the ship channel was calculated. Because there are no major airports in the vicinity of Matagorda Bay, aircraft accident statistics for general aviation were used in these calculations. The main contributing factor to aircraft–ship accidents in the ship channel is the operation of small planes at the local Port O'Connor airstrip.

There is also a helicopter landing site in Port O'Connor, from which approximately 25 flights leave daily for oil platforms in the Gulf of Mexico. Three companies operate these flights. Accident possibilities from these flights are not included in the calculation, as it is assumed that flight patterns can be arranged to avoid operations in the area where the LNG carrier will operate.

If an airplane crashes into the storage tanks or LNG pipelines, for reasons discussed earlier it is assumed that there is no risk to the public. For crashes into ships, it is assumed that the spill would be 10,000 cubic meters in 12 minutes.

The potential risks due to sabotage are qualitatively examined. It would appear essentially impossible to eliminate the likelihood of sabotage by determined saboteurs. However, it is very likely that immediate ignition would occur due to the violence required to release the LNG in the first place. In the broader sense, the decision to allow LNG import terminals anywhere in the US implies that such risks must be taken.

Onshore and marine terminal accidents. Wesson and Associates, Inc. (WA) conducted an analysis of the potential consequences resulting from accidental LNG spill events in the Onshore and Marine Terminal Facilities. Specifically, the downwind vapor dispersion characteristics and thermal radiation levels were calculated for postulated spill events including maximum credible spills on the pier, spills in the unloading and sendout transfer piping systems, and spills from the LNG storage tanks and associated piping.

The postulated LNG storage tank spill is impounded in a diked area surrounding each LNG tank. LNG spill collection troughs, run-

off trenches and remote impounding basins are provided for all other spill events. This assures that any accidental release of LNG and its subsequent spreading is contained at all times within the plant boundaries.

All spill impounding areas, including the LNG storage tank diked areas, are provided with automatic vapor dispersion and fire control systems. These systems insure that no thermal radiation or vapor dispersion hazards exist at any plant boundary line under any set of weather conditions.

Marine accidents. Engineering Computer Opteconomics, Inc. (ECO) conducted an analysis of ship accidents leading to an LNG cargo release. Specifically, the probabilities of collision with other vessels, rammings of non-vessel, stationary, or floating objects, and grounding were calculated for three subsets within the Matagorda Bay port scenario:

- the entranceway, which is the immediate ship approach route to and through the cut in Matagorda Peninsula;
- the harbor, which is the ship channel within Matagorda Bay; and
- the piers, which are the waters between the ship channel and the berth.

The analysis of spills due to collisions indicates that either one or two cargo tanks may be involved. A one tank cargo involvement is considered the *most credible spill event* due to collision, and would involve 19,400 cubic meters of LNG. The *maximum credible spill event* due to collision is considered to be the simultaneous rupture of two adjacent cargo tanks, and could involve 38,800 cubic meters of LNG. The analysis of spills due to ramming indicated that the *most credible spill due to ramming* is 10,000 cubic meters. The analysis of grounding indicated that the expected value for cargo tank rupture is so small as to be statistically insignificant; thus they do not pose any significant potential for cargo release and are not considered further.

Summary of initiating events. The quantified initiating accidents and the accompanying spills that were further examined to determine public risk are summarized in Table 1.

Vapor cloud formation and downwind travel

If a spill of sufficient size occurs without immediate ignition, an LNG vapor cloud may form and travel downwind. The characteristics of the cloud depend for the most part on the size and rate of the spill (Table 1) and the surface area on which the LNG is spilled. Since land spills have been ruled out as significant to public risk, the focus here is on water spills. The travel of the vapor cloud depends on the wind speed and direction and the stability of the weather.

Likelihood of immediate ignition. The potential for a fire associated with the releases depends on both the presence of flammable vapors and an ignition source. Vaporization would commence immediately following a cargo tank penetration. Ignition of vapor can be caused by any heat source of sufficient temperature and duration. The primary ignition sources would be the friction and sparking generated by the immense forces needed to attain penetration. Temperatures can be expected to range from 1600°F to 2700°F, far in excess of the 1000°F ignition temperature of methane and air. Secondary ignition sources such as boilers, galley fires, electrical cables, light fixtures, etc. will also be present and exposed as a result of collisions, rammings, or aircraft/ship crashes. Therefore it is extremely likely that immediate ignition will occur under such circumstances. In the analysis it is assumed that the probability of ignition following such a penetration is 0.99.

Wind direction and speed and weather stability. Because wind direction, wind speed, and stability class are not independent of each other; the weather conditions in the vicinity of the La Salle Terminal have been tabulated as illustrated in Table 2 and more details were available for various wind speeds and stability classes [2, 3]. In compiling the tables, stability classes *A* and *B* were assumed to be *C* and class *E* was assumed to be *F*. Also, winds of 20 *mph* and greater were assumed to be 15 *mph*, which results in greater plume lengths. All these assumptions are conservative. Notice that stability class *F* does not occur with a 15 *mph* wind.

For the most credible spill of one tank (19,400 cubic meters spilled instantaneously), with a 10 *mph* wind and stability class *D*, the vapor cloud could travel up to 5.88 miles if it is not ignited. This would require roughly 25 minutes. Then the cloud would subside as

Vol. 7
1979

TABLE 1. QUANTIFIED ACCIDENTS CONTRIBUTING TO PUBLIC RISK

Location of accident	Cause and size of spill	Annual probability of accident event	Spill size
Entrance	Most credible spill due to collision	3.46×10^{-5}	19,400 m^3 instantaneously
	Maximum credible spill due to collision	9.28×10^{-6}	38,800 m^3 in 9 min
	Spill due to ramming	7.59×10^{-8}	10,000 m^3 in 12 min
	Spill due to aircraft crash	1.21×10^{-8}	10,000 m^3 in 12 min
Harbor	Most credible spill due to collision	6.49×10^{-7}	19,400 m^3 instantaneously
	Maximum credible spill due to collision	1.76×10^{-7}	38,800 m^3 in 9 min
	Spill due to ramming	2.10×10^{-7}	10,000 m^3 in 12 min
	Spill due to aircraft crash	9.23×10^{-8}	10,000 m^3 in 12 min
Pier	Most credible spill due to collision	1.01×10^{-7}	19,400 m^3 instantaneously
	Maximum credible spill due to collision	2.73×10^{-8}	38,800 m^3 in 9 min
	Spill due to ramming	1.10×10^{-5}	10,000 m^3 in 12 min
	Spill due to aircraft crash	5.93×10^{-7}	10,000 m^3 in 12 min

the LNG-air mixture was reduced to below the lower flammable limit of 5% average concentration.

Probability of size and location of cloud at ignition and the extent of its effects

This requires identification of the potential ignition sources and the development of a vapor cloud ignition model and a thermal radiation model.

Identification of potential ignition sources. A vapor plume can be ignited from numerous sources such as spark plugs, open flames, pilot lights, electrical sparks due to short circuits, and so on. Because of the difficulty of tabulating all possible ignition sources, it is conservatively assumed that each house or building has only one source and that each recreational boat or commercial fishing ship has one source. To assume fewer sources is conservative because it implies increased likelihood that a vapor cloud would cover a larger area before ignition.

TABLE 2. PROBABILITIES OF WIND DIRECTIONS NEAR LA SALLE TERMINAL

Wind direction	Probability of occurrence
N	0.123
NE	0.103
E	0.131
SE	0.338
S	0.196
SW	0.037
W	0.020
NW	0.052

Vapor cloud ignition model. The structure of the ignition model can be illustrated as follows. The probability that any individual source ignites a passing vapor plume is defined to be p, and this is identical for each ignition source. Hence, if a plume begins to cover an area with ignition sources, the probability it ignites from the second source is $(1 - p)(p)$, which is the probability $(1 - p)$ that it does not ignite at the first source times the probability (p) that it does ignite at the second source. In general, the probability that a vapor plume is ignited by the n^{th} source is $(1 - p)^{n-1}(p)$.

Radiation model. In this study, a thermal radiation level of 5300 Btu/ft^2/hr is assumed to be the lower limit resulting in fatalities. This is conservative because one would actually expect only blistering of skin exposed for five seconds to such thermal radiation. Using such a level results in the distances calculated in Table 3 for LNG pool and plume fires. Because these do not vary greatly with weather conditions, these distances are assumed for all weather conditions.

TABLE 3. DISTANCES ASSUMED FOR THERMAL-RADIATION FATALITIES

Incident	Distance from 5300 Btu/ft²/hr exposure for 5 secs
10,000 m³ in 12 min pool fire	1100 feet from spill center
19,400 m³ instantaneous spill pool fire	2500 feet from spill center
38.800 m³ in 9 min pool fire	2500 feet from spill center
plume fire	525 feet from plume fire

IV. RISK QUANTIFICATION

The purposes of the risk quantification were to determine (1) the expected number of public fatalities per year due to the LNG terminal operation, (2) annual risk levels (i.e. the probability of becoming a fatality) to individuals who might possibly be affected by the facility, and (3) the probability that different fatality levels might be exceeded in any year.

This required, in addition to the information previously developed, data on the population distribution in the vicinity of the La Salle Terminal.

Population study

A detailed on-site study of the population living in and visiting the vicinity of the La Salle Terminal was conducted. The main goal was to determine how many people are at which locations, at which times. Because of large seasonal and weekly fluctuations in the population, population data were tabulated separately for all combinations of days and nights, weekdays and weekends, tourist season (April through October), and nontourist season. After a preliminary examination of population data, it was decided that three combinations—nontourist season, tourist weekends, and tourist weekdays—adequately described population distributions. The risk quantification considered each of these cases separately.

Data used to estimate populations included identifying houses as belonging to permanent residents or transient visitors through electrical use information, current and projected school enrollments, numbers of boats, estimated beach visitors, fishing season dates, special events in the area, etc. For fatality calculations, the population study indicated that it was reasonable to assume 3.62 people per permanent household, 4.02 people per occupied transient household, 4.5 people per recreational boat, and 2.5 people per commercial fishing ship.

Illustrative calculation

To indicate the structure of the risk analysis model, the contribution to the average number of fatalities made by one accident scenario is illustrated. Define accident scenario S_1 as a collision between an LNG carrier and another ship in the harbor at the intersection of the ship channel and intercoastal waterway. This collision produces an instantaneous one-tank spill of 19,400 cubic meters of LNG. Furthermore, assume the spill does not immediately ignite, so that a vapor cloud forms. The wind is assumed to be from the east at 10 mph with a Pasquill stability class *D*. The eighth ignition source ignites the vapor cloud. The illustration is for a weekday during the tourist season.

The probability of this accident scenario is the product of the following factors:

- the annual probability of a collision releasing one tank of LNG (most credible spill) in the harbor during a weekday in tourist season (Table 1)[3];
- the probability $p = 0.01$ of no ignition given this collision;
- the probability that the wind is from the east (Table 2);
- the probability that given an east wind, it is 10 mph and there is stability class *D* and;
- the probability that the eighth ignition source ignites the vapor cloud.

The calculation, using the probabilities from the tables indicated is as follows:

$$Pr(S_1) = (2.33 \times 10^{-7})(0.01)(0.131) \times (0.1856)(0.0478) = 2.71 \times 10^{-12}. \quad (8)$$

[3] The probabilities presented in Table 1 were multiplied by the ratio of each time period (nontourist season, tourist weekends and tourist weekdays) in days to 365 to give the probability of an accident within a time period.

Vol. 7
1979

To calculate the expected fatalities $F(S_1)$ from accident scenario S_1 on a tourist season weekday, the maximum extent of the flammable vapor plume if it is not ignited is first determined.

Since 100 recreational boats are assumed to be in the designated region of Fig. 1 and, since the projected plume covers two percent of the region, the expected number of recreational boats in the plume's path is 2, assuming a uniform distribution of boats. No commercial boats are assumed in this area on tourist season weekdays.

On summer weekdays, there is an average of 1000 daytime transient visitors in Port O'Connor. It is conservatively assumed that all of these visitors are on the beach on the east side of Port O'Connor, so that with 34% of the beach covered by the plume, it is assumed that 340 daytime transient visitors are on the beach within the vapor cloud.

The number of houses occupied permanently and covered by the maximum possible extent of plume is 37, and the number of houses occupied by transients is 103 for a total of 140 household ignition sources. The average weekday daytime occupancy of these houses is taken as 2.62 and 3.02, respectively. This assumes one person is away from each household. Averaging the expected weekday-daytime occupancy of the houses occupied permanently and by transients, an expected number of people per household source is determined to be 2.91.

In the event of a house-caused ignition, it is assumed that individuals in boats and transient daytime visitors within the vapor cloud, as well as all individuals in houses covered by the cloud at the time of ignition, are fatalities.

The calculation of the contribution of accident scenario S_1 to expected fatalities is as follows. Since the vapor cloud is assumed to be ignited by the eighth ignition source, it is not ignited by the two boats and it is ignited by the sixth house encountered. The number of individuals within this cloud at the time of ignition is the people in the two boats, on the beach, and in the six houses under the cloud. Summing these gives the expected number of fatalities in the cloud as

$$2(4.5) + 340 + 6(2.91) = 366.46. \quad (9)$$

As discussed, it is assumed that individuals within 525 feet of a vapor cloud fire would be fatalities due to thermal radiation. Using the average population density of Port O'Connor, it is assumed that 230 individuals could be within 525 feet of a plume fire. 20% of these are assumed outside at any given time. Hence, the total expected fatalities from S_1 on a summer weekday is

$$F(S_1) = 366.46 + (0.2)(230) = 412.46. \quad (10)$$

Multiplying (8) times (10) results in 1.118×10^{-9}, which is the contribution $F(S_1)$ of accident scenario S_1 to the overall annual expected fatality rate F in (3).

Calculations of expected fatalities due to thermal radiation from an LNG pool fire are illustrated by the following example. The probability of an LNG carrier having a collision in the harbor during a tourist season weekend is 1.20×10^{-7} per year. Based on data from the Matagorda Bay area, the average number of boats within 2500 feet of the center of that spill is 20. The average number of occupants in a recreational boat is assumed to be 4.5 and the average number of occupants in a small commercial fishing boat is assumed to be 2.5. Since slightly less than 75% of the total number of boats exposed to pool fires are recreational boats, an average occupancy per boat of 4.0 is used. The annual expected fatalities F_f due to a pool fire resulting from a collision in the harbor on a tourist season weekend is

$$F_f = (1.20 \times 10^{-7})(0.99)(20)(4) = 9.50 \times 10^{-6}. \quad (11)$$

Results

The annual risks for individuals in different groups are presented in Table 4. The probability of exceeding a specified number of fatalities in a given year is shown in Fig. 4.

The relative contribution to the annual expected fatalities of different accidents, seasons, locations, and spills was determined. These are shown in Table 5. As shown in Table 5(c), approximately 62% of the expected fatalities come primarily from individuals in boats subjected to the pool fire thermal radiation resulting from immediate ignition of an LNG spill in the harbor during tourist season weekends. The balance of the expected fatalities, 38%, is from the delayed ignition of the vapor clouds. Approximately half of this 38% is from

TABLE 4. SUMMARY OF PUBLIC RISKS

Group	Expected fatalities per year	Number of people in the group sharing the risk	Risk per person per year
Permanent population in Port O'Connor	2.0×10^{-8}	800	2.5×10^{-11}
Permanent population in Indianola	1.3×10^{-7}	80	1.7×10^{-9}
Transient daytime visitors	2.5×10^{-6}	2500	9.9×10^{-10}
Individuals in boats	1.35×10^{-5}	3000	4.5×10^{-9}
All individuals exposed to risk	1.7×10^{-5}	9000	1.9×10^{-9}

individuals in boats, and half from individuals on shore.

Sensitivity analyses were conducted to examine the effect of variations on two parameters on public risk. The two parameters were the probability of ignition per source and the average methane concentration at which the LNG vapor cloud is still ignitable. The base case assumed that the probability of ignition per source was 0.1 and that an LNG vapor cloud was ignitable until it deteriorated to 5% average methane concentration. These two parameters were varied, one at a time, to generate two additional, more conservative cases:

(i) *Probability of ignition per source of 0.01 and average concentration condition.* This case examined the effect of a lower ignition probability on public fatalities. With the lower probability of ignition, the expected number of boats and/or houses covered by an LNG vapor cloud at the time of ignition would increase; hence the expected fatalities would be higher. This case, therefore, analyzes a more conservative assumption with regard to ignition probability.

(ii) *Probability of ignition per source of 0.1 and peak concentration condition.* There is disagreement among experts with regard to the distance that ignitable (i.e. within the 5–15% flammable concentration range) portions of an LNG vapor cloud would travel. It is hypothesized that pockets of this concentration may exist even after the average concentration of the cloud has decreased considerably below the 5% average methane concentration. To address this possibility, a conservative condition is analyzed. It was assumed that pockets of peak concentrations that are still ignitable exist in the cloud until it reaches an average concentration of 2.5%. Plumes whose boundaries were determined by 2.5% average concentration under different conditions of wind speed, stability class, and spill size were obtained from Wesson and Associates. These plumes were used to calculate exposed populations in the same way the 5% average concentration plumes were used.

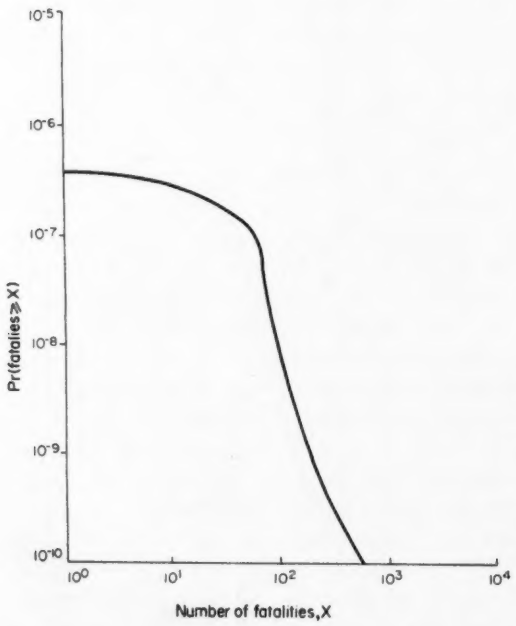


FIG. 4. *Annual probability of exceeding a specified number of fatalities.*

Both cases showed a slight increase in the expected fatalities per person per year over those for the base case. For the two cases outlined above, the expected fatalities per person per year were 2.22×10^{-9} and 2.21×10^{-9} respectively. Comparison with the expected fatalities of 1.9×10^{-9} per person per year for the base case shows that the public risks are not influenced significantly by more conservative estimates of the two parameters evaluated in the sensitivity analysis.

Vol. 7
1979

TABLE 5(a). EXPECTED FATALITIES FOR DIFFERENT SEASONS

Season	Expected fatalities per year	Percentage of the total expected fatalities
Nontourist (November through April)	8.21×10^{-7}	4.8%
Tourist Weekdays (May through October)	3.26×10^{-6}	19.2%
Tourist Weekends (May through October)	1.29×10^{-5}	76.0%

TABLE 5(b). EXPECTED FATALITIES FOR DIFFERENT ACCIDENT LOCATIONS

Accident location	Expected fatalities per year	Percentage of the total expected fatalities
Entranceway	2.45×10^{-6}	14.4%
Harbor	1.10×10^{-5}	64.8%
Pier	3.52×10^{-6}	20.8%

TABLE 5(c). EXPECTED FATALITIES FOR DIFFERENT LNG SPILLS*

Type of LNG spill	Expected fatalities per year	Percentage of the total expected fatalities	
Most credible spill due to collision	9.70×10^{-6} (2.25×10^{-6})	57.2%	(13.3%)
Maximum credible spill due to collision	2.72×10^{-6} (6.9×10^{-7})	16.0%	(4.1%)
Spill due to ramming	3.63×10^{-6} (3.15×10^{-6})	21.4%	(18.6%)
Spill due to aircraft crash	8.99×10^{-7} (2.75×10^{-7})	5.4%	(1.6%)
Spill resulting in pool fire	(1.06×10^{-5})		(62.4%)

* Numbers in parentheses are obtained by separating expected fatalities due to plume fire and pool fire.

V. RISK EVALUATION

The La Salle Terminal is designed to receive and vaporize approximately one billion cubic feet of natural gas per day. For risk comparison, the energy output of this natural gas is equivalent to about eighteen 1000 megawatt electric power plants operating at 70% capacity factor [8].

Buehring [4] has calculated the expected fatalities per year due to 1000 megawatt coal power plants. Based on 1970 data for the State of Wisconsin, the expected number of deaths to the public due to transporting coal was 0.695 per year for a 1000 megawatt coal facility. Multiplying by 18, this becomes 12.51 expected fatalities per year. For the equivalent energy output of La Salle Terminal, the expected fatality level is 0.000017. This is less than one hundred-thousandth (1/100,000) of the expected fatalities from the equivalent consumer energy source from coal.

To compare the individual risk levels, causes of death and their risk per individual per year are given in Table 6 (US Department of Commerce, 1975). The risk to an average individual in the US due to fire is 16,000 times greater than the risk to an individual exposed to the operations of the proposed La Salle Terminal.

The group with the highest annual risk of 4.5×10^{-9} per year is composed of the individuals in boats. This is one chance in approximately 220 million. For Port O'Connor, the annual risk per person is 2.5×10^{-11}. The National Safety Council [5] provides data to calculate that the expected public risk due to gas distribution systems in the U.S. is 5.15×10^{-7} per individual per year, which represents one chance in 1.9 million. Hence, for any individual exposed to a hazard because of the project, the probability that he will die due to a natural gas distribution accident is 271 times as great as the possibility due to the operations of the proposed La Salle Terminal.

The data provided by the National Safety

TABLE 6. SOME MAJOR CAUSES OF DEATH—1973

Cause of death	Deaths	Risk per person per year*
Motor vehicle accidents	55,511	2.65×10^{-4}
Suicide	25,100	1.20×10^{-4}
Murder victims	17,123	8.10×10^{-5}
Fire and flames	6503	3.10×10^{-5}
Firearms	2618	1.25×10^{-5}
Aircraft accidents	1638	7.81×10^{-6}
Electricity current in the home	232†	1.11×10^{-6}
Gas distributed by pipeline	108†	5.15×10^{-7}

* 1973 population used was 209,860,000.
† National Safety Council, 1975.

Council [5] on the risks due to electric shock in electrically wired residences show that expected public fatalities due to this risk are 1.11×10^{-6} per individual per year. This represents one chance in 0.9 million. Hence, in an electrically wired house, the probability that the individual will die next year due to electric shock in his home is 584 times as great as the possibility of death due to the operations of the proposed La Salle Terminal.

For the permanent population of Port O'Connor, the same model used to calculate the probability of airplane crashes into LNG carriers was used to calculate the probability that an airplane would crash directly into Port O'Connor and kill an individual. Using the angular dependence model, this probability was found to be 8.60×10^{-8} per year for a permanent resident living one mile in front of the runway. Using the aircraft crash model with no angular dependency, the probability of an individual living a mile from the runway becoming a fatality is calculated to be 4.73×10^{-8} per year. About 600 of the 840 permanent residents of Port O'Connor live within one mile of the runway. The individual risk due to air crashes in Port O'Connor is over three orders of magnitudes higher than the risk to a Port O'Connor resident due to the operation of the proposed terminal.

Approximately 65 meteorites weighing more than one pound hit the United States each year. If one owns a one-floor house with 3050 square feet, the probability that one of these meteorites will hit that house within a year is 1.9×10^{-9}. This is identical to the average individual risk due to operation of the La Salle Terminal.

A criterion for risk acceptability has been proposed by Starr [6]. It suggests that acceptability be determined by the annual probability of death per person exposed and the benefits of the activity, measured in arbitrary units. Because the evaluation of benefits is complex, the main use of Starr's criterion is in establishing a bound of approximately 10^{-7} for acceptable risk levels. The individual risk from the operations of the proposed La Salle Terminal is 1.9×10^{-9}, which is less than 2% of 10^{-7}. The societal risk due to operation of the La Salle Terminal is, consequently, much less than Starr's criterion for social acceptance.

VI. CONCLUSIONS

The formal risk analysis described in the previous sections presented several advantages:

- data and professional judgments from experts in various fields could be integrated into a logical framework;
- assumptions could be stated explicitly;
- sensitivity analyses could be conducted to evaluate the significance of the assumptions;
- the public risk could be systematically estimated and approved; and
- strategies to reduce risks could be developed

The results of the analysis showed that the risks to the public from the operation of the La Salle Terminal were below those that the population in the vicinity of the terminal is exposed to at the present time. The risk levels

are also less than those published in literature as acceptable to society.

The study was used in preparing the Safety Analysis report for submission to the Federal Power Commission (FPC). The Final Environmental Impact Statement issued by the FPC has stated that the levels of public risk associated with the La Salle Terminal facility are acceptable.

ACKNOWLEDGEMENTS

This study was conducted for the El Paso LNG Company and its subsidiaries. Mr. Ivan Schmitt, Vice President, was responsible for the Safety Analysis report. Mr Randall I Cole assembled the team of consultants and was a constant source of advice and contributions. Mr Joe Porricelli (ECO) provided all the ship accident and spill information, and Dr. Hal Wesson provided input about LNG operations and the vapor cloud and ignition models. Their experience and knowledge were critical to the study. Other members of Woodward-Clyde staff who made significant contributions to the study were Steven James, Craig Kirkwood, John Pelka, and John McMorran.

REFERENCES

1. El Paso Atlantic Company, El Paso Eastern Company, El Paso LNG Terminal Company (1977). *Joint LNG safety report respecting the proposed Algeria II project before the Federal Power Commission*, Docket Nos. CP 73–258 *et al.*, Volume II of II.
2. GLENN AH and Associates (1976) *Meteorological–Oceanographic Conditions Affecting Design and Operations of Proposed LNG Tanker Terminal Facilities, La Salle Terminal, Southwest Matagorda*, Texas, New Orleans, Louisiana (see also Appendix A of [1]).
3. National Oceanic and Atmospheric Administration (1971) *Wind distribution by Pasquill stability class (star program) seasonal and annual*. National Climatic Center Federal Building, Asheville, North Carolina, 6/29/71 (see also Appendix B of [1]).
4. BUEHRING WA (1975) *A model of environmental impacts from electrical generation in Wisconsin*. PhD dissertation, Department of Nuclear Engineering, University of Wisconsin, Madison.
5. National Safety Council (1975) *Accident Facts*, 1975 *Edition*, Chicago, Illinois.
6. STARR C (1969) Social benefit versus technological risk. *Science N.Y.*, 1965 p. 1232.
7. US Bureau of the Census (1975) *Statistical Abstract of the United States; 1975*, 96th Ed. Washington, DC.
8. VAN HORN AJ & WILSON R (1976) Liquefied Natural Gas: Safety Issues, Public Concerns, and Decision Making, Energy and Environmental Policy Center, Harvard University, Cambridge, Massachusetts.

ADDRESS FOR CORRESPONDENCE: *Dr Ralph L Keeney, Woodward–Clyde Consultants, Three Embarcadero Center, Suite 700, San Francisco, California* 94111, *USA*.

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 3. pp. 207 to 218
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0601-0207$02.00/0

A Supportive Application of Management Science to Marketing Strategy Formulation[1]

JEAN-CLAUDE LARRÉCHÉ

INSEAD, Fontainebleau, France

EDWARD C STRONG

Tulane University, New Orleans, Louisiana

(*Received June* 1978)

Traditional applications of management science aiming at solving specific decision-making problems are claimed to be often of an *aggressive* nature. They represent an intrusion into the life space of the managers who are supposed to use them and in many cases are consequently rejected. In the light of the evolution of marketing and corporate modelling over the last twenty years, the authors propose a new *supportive* approach for management science applications which is user-oriented rather than task-oriented. The characteristics of the aggressive and supportive approaches are contrasted, and an example of a supportive management science application for marketing strategy formulation is described.

SINCE THE widely publicized BBDO–CEIR media selection model [5], a large number of marketing models have been developed and described in the literature [13, 26, 30]. When examining this literature, two observations come readily to mind. Firstly, the actual use of these models in business remains very limited despite a substantial volume of research. Secondly, most applications of management science to marketing are concerned with tactical decisions such as media selection [22], sales force allocation [25] or competitive bidding [8]. There is a scarcity of management science applications to unstructured or strategic marketing decisions such as product portfolio management or the formulation of product market strategies.

The purpose of the present paper is to discuss a neglected class of management science applications which offer some definite advantages in terms of the two dimensions mentioned above: successful implementation and contribution to unstructured strategic issues.

These applications aim at developing the professional competence of managers rather than solving specific managerial problems and they include pedagogical games, as well as computer-assisted instruction packages. They are termed supportive because their primary purpose is to help the manager to improve his performance, without interfering with the decision-making process. It will be argued that these supportive applications represent an important untapped potential for management science.

In the following sections of this paper, the evolution of management science applications in marketing and in corporate modeling is reviewed in order to extract the main trends in the implementation of these applications. In the light of these emerging trends, the supportive approach is compared with the traditional management science approach. Finally, an example of a supportive management science application in the area of marketing strategy is described.

[1] The first version of this paper was presented at the TIMS International Meeting in Athens, Greece, July 25–27, 1977.

THE EVOLUTION OF MANAGEMENT SCIENCE APPLICATIONS IN MARKETING AND CORPORATE PLANNING

Comparing the evolution of management science applications in marketing and corporate planning is of interest for two main reasons. First, the design of applications in marketing strategy should obviously benefit from past experience in corporate modeling as these two areas tackle some common substantive and often ill-structured issues. Secondly, management science applications have evolved practically independently in these two areas. Moreover, these two areas have different organizational properties. Marketing tends to represent a larger portion of the organization with staff and line managers from different backgrounds and with a wide variety of educational levels. Corporate planning is a staff organization usually composed of a limited number of managers from different functions but with relatively high educational levels. It is thus valuable to investigate whether these differences have resulted in a different evolution of management science applications in these two areas.

Marketing models

There has been a proliferation of marketing models in the last fifteen years. Some of these models were developed for managerial applications, others represented research efforts to improve the managerial validity of some basic techniques by relaxing some of the more constraining assumptions as, for instance, in the area of stochastic modeling. Montgomery and Weinberg [27] claim that "the vast majority of really managerially useful (marketing) modeling work is past 1967", and they distinguish between three generations of marketing models:

1. Emulation of classical operations research (1959–1964). "The earliest era of marketing models was characterized by techniques in search of problem". The formulation of marketing problems was oversimplified in order to solve them through classical operations research optimization techniques. As a result, the recommendations provided by these models were of little managerial value. A typical example of this generation is the media selection model developed by BBDO–CEIR [5] and based on linear programming;

2. Bigger and better models (1964–1970). The reaction to the lack of success of previous management science applications to marketing was to develop more complete models which could encompass more of the complexity inherent to marketing phenomena. The extreme example of this generation is certainly the simulation developed by Amstutz [2] and applied to the pharmaceutical industry [6]. Although these applications have generally been more successful than the previous ones, they tended to be too complex for management understanding, continuing use, and control;

3. The decision calculus (1970–). In 1970, Little [19] called 'Decision Calculus' a model-building approach integrating data and managerial judgements and placing an emphasis on the effective use of the model by managers. Little specifies that a decision-calculus type of model should be simple, robust, easy to control, adaptive, complete on important issues, and easy to communicate with. In particular, the completeness/complexity dilemma of the two previous generations has been resolved by the incorporation of managerial subjective judgements which integrate the numerous interrelationships present in marketing phenomena, while keeping the model structure relatively simple. These models appear to be more successful than the previous ones in terms of number of applications [1]. The typical examples of decision-calculus models are ADBUDG [19] and BRANDAID [20, 21] but several models have been developed along these lines since 1970 (e.g. [3, 23, 25]).

The above historical perspective clearly shows a slow and expensive trial-and-error process of searching for new implementation approaches in order to improve the acceptance of marketing models by managers.

This experience seems to direct management scientists towards simpler models in which complexity and technical sophistication are traded for ease of understanding, use and control. This is confirmed in a study by Larréché and Montgomery [15] where more recent models of the decision-calculus type are rated

Vol. 7
1979

more highly on implementation-related dimensions by a group of experts than more traditional models.

In addition, many applications [1, 19, 23, 27] indicate that the pedagogical impact of marketing models on managers may be as important as direct improvement of specific decisions. A modeling application provides marketing managers in particular an opportunity to make explicit and reconsider their decision-making processes, a formal structure for reviewing the major factors influencing a given situation, a framework for planning data-gathering activities, and a basis for effective communication. As a result, some models such as ADBUDG [19] or SPRINTER I [33] have sometimes been used as pedagogical devices prior to the implementation of more complex models [21, 25, 32].

Corporate models

Corporate models seem to have known a much higher degree of commercialization than marketing models. Special purpose computer languages such as IFPS [29], PROJECTOR [24], PROPHI II and SIMPLAN [28] have also been developed and distributed. Several surveys [4, 9, 10, 28] indicate that a large and growing number of companies make use of corporate models. Naylor and Schauland report: "Through direct personal contact we have identified over 2,000 corporations in the United States, Canada, and Europe who are either using, developing or planning to develop some form of corporate planning model" [28]. The following points also emerge from these surveys which were undertaken in North America and Europe:

Most corporate planning systems are finance-oriented. They primarily do cash-flow projections and the marketing component of the model is at best an econometric formulation relating sales to marketing instruments and exogenous variables. It is unlikely that these models may efficiently be used to test the effects of major changes in product-market strategies.

Most applications use simulation to answer 'what if?' types of questions, rather than optimization techniques as in the earlier corporate models. There are still some advocates of mathematical programming techniques [12, 34] but they seem to concentrate in the oil industry where this tradition has been established in operational planning, and they seem to be stronger in Great Britain than in North America, an observation also reported by Boulden [4].

The vast majority of the corporate models use a deterministic approach rather than the more complex stochastic or risk analysis approaches.

These observations clearly indicate that most of the corporate models in use are relatively simple in terms of their scope and of the management science techniques employed. This is also true in terms of their costs: in a survey of 65 British companies, Grinyer and Wooller [10] found that the average reported total cost to develop a corporate model was below $20,000.

The characteristics of the typical corporate models currently used thus appear to be very similar to those of the decision-calculus models in marketing. However, according to an article by Hayes and Nolan [11], these similar current characteristics of marketing and corporate models are the result of similar trial-and-error developments. Hayes and Nolan distinguish three phases in the history of corporate models, corresponding to the bottom-up, top-down, and inside-out approaches. As indicated in Table 1, the characteristics of these phases, as well as their dates, are analogous to those of the three generations of marketing models, as described by Montgomery and Weinberg [27].

Hayes and Nolan also stress the pedagogical role of corporate models: "a good corporate model is not only a decision-making aid, but also a powerful educating and developing tool for management" [11, p. 110]. Finally, Grinyer and Wooller [10] as well as Naylor and Schauland [28] have found that acceptance of corporate models depended largely on their flexibility when used by managers, a dimension also underlying the decision-calculus concept in marketing modeling.

Emerging trends in management science applications

From the above cursory review of developments in marketing and corporate modeling, it is clear that, despite obvious differences in the substance and the organizational context, these two areas have evolved in a similar fashion over the last 20 years. A number of lessons have been learned which revolve around the following three points:

1. Greater simplicity and flexibility of the models;
2. The consideration of users' characteristics,

TABLE 1. THE EVOLUTION OF MARKETING AND CORPORATE MODELS

Marketing models[1]	Corporate models[2]
Emulation of classical OR (1950–1964) "Techniques in search of problems"	*Bottom-up approach* (1956–1963) Corporate models designed and implemented by technically-oriented personnel, from operating models.
Bigger and Better models (1964–1970) More complete models, too complex for management understanding, continuing use and control.	*Top-down approach* (1964–1969) Large-scale models beyond the comprehension of managers
Decision calculus (1970–) Emphasis placed on the effective use of simpler models integrating data and managerial judgments.	*Inside-out approach* (1970–) Managers intimately involved in the development of simple models.

[1] As described in Montgomery and Weinberg [27].
[2] As described in Hayes and Nolan [11].

or the involvement of users, in the design of the models;

3. Recognition of the pedagogical value of models and of the modeling process.

A NEW LOOK AT MANAGEMENT SCIENCE APPLICATIONS: AGGRESSIVE VERSUS SUPPORTIVE APPROACHES

Important progress has been made in problem solving since the Second World War, thanks to management science. Processes and techniques have been developed, and diffused, which limit the risk of major conceptual and technical mistakes in management science applications. Like the experienced craftsman, management scientists can be proud of the quality of the products which they turn out. This traditional approach rests, however, on the assumption that a product which is technically good, which adequately solves a given problem, will automatically be used. It is backed up by popular wisdom reflected in the English saying, 'Build a better mousetrap and the world will beat a path to your door'. This does not always work, as many management scientists and marketers have unfortunately experienced.

Models are not developed in a show-case with managers eagerly waiting for them to be available. Managers have their own objectives, activities, environment, as members of an organization, as members of a profession, and as human beings. To use Lewin's terminology [18] they have a life space of their own which has been determined through their personal and professional history. The introduction of a model may be considered as the transplant of a foreign object into their life space. This transplant may be rejected, even though it is of good technical quality, if it does not fit the needs of the managers who are supposed to use the model. These needs have been defined here in a general sense in terms of the individual's life space and not only in terms of the specific problem to be solved.

There are many cases of rejections of 'good' models by managers, and Montgomery and Weinberg [27] report an excellent example of this phenomenon. It is the case of a competitive bidding model, slightly more sophisticated than the RCA one [8]. Competitive bidding models provide good examples as their 'goodness' may easily be measured by the percentage of bids won over a period of time. This model was used because a particularly skilled operations research manager assisted the managers in the process of providing judgements to the model. Using this model, the company won twelve out of fourteen bids, so there was no question about the technical validity of the model, and this was recognized easily by the managers. However, "when the operations research manager became busy with other activities and the managers were left on their own to use the model, they reverted to the previous,

Vol. 7
1979

less-disciplined bidding process. The company then lost three of the next four bids while not using the model. The apparent reason that managers stopped using this successful model was that the old fuzzy format was more comfortable and, after the OR manager went on to other things, they found it easier to go back to the old way ... the model has not been used since then, despite its effectiveness" [27, p. 26].

The traditional management science approach is essentially task-oriented. From a technical point of view, the tools developed may be valid as they may indeed contribute to improving the performance of specific managerial tasks, at least under the assumptions made in the modeling process. But these new tools will require changes in the life space of the managers, changes which are often unforeseen, not prepared for, and not commensurate with the personal rewards that they will produce. This is why we have labelled this traditional approach the aggressive approach. In a business context, it would be called a 'production' or 'selling' orientation as opposed to a marketing one.

All management science applications following the traditional approach are not doomed to failure, and some of them eventually succeed when the conditions for a successful transplant happen to be there. The traditional approach also has its merits and may have been historically necessary given the fact that the bottleneck in the management science process has been at the technical level for quite a while. But the long-term survival of management science in the business context is highly uncertain if it keeps this production or selling orientation.

Our contention is that there is an alternative to the traditional approach which would be user-oriented rather than task-oriented, which would be marketing-oriented rather than production or selling-oriented. We have labelled it a 'supportive approach'. The supportive approach recognizes that the needs of the user should not be defined uniquely in terms of the specific problem at hand but in a much broader sense in terms of his life space, including his personal objectives, organizational objectives, immediate environment, time constraints and cognitive constraints. For instance, an interesting observation has been made by Simon, Lamar and Haines [31] concerning the use of models by managers. While most models aim at solving an operational problem, or supplying managers with new information, these authors claim that "such models have only limited import relative to the needs of management of any large organization." Instead, they observe that three other types of model uses are made by managers: to force one part of an organization which has data to give it to another part of the organization which wants it; to provide a rationale for decisions previously reached by the managers based upon informal analysis; and to establish the manager's reputation. These uses certainly correspond to broader needs than the problem-solving ones for which models are usually developed.

The foundation of the supportive approach is thus a user-orientation as opposed to the more traditional task-orientation. 'Users' are the managers who are in a position to exploit the benefits of a potential application. They are the people who are going to be affected by the application and who are going to influence most directly its long-term success in the organization.[2]

This user-orientation implies a number of propositions for the implementation of management science. The main ones are summarized and contrasted to those of the task-orientation in Table 2. The first one concerns the objectives of management science. While within a task-orientation the primary objective is to improve the performance of specific management task, in a user-orientation it is to satisfy users' needs. These needs may themselves be defined in terms of task performance, but they will usually be broader. They may for instance include the desire to increase one's competence, to build up confidence in one's decisions, to have a means to focus on major potential problems, or to prepare oneself for

[2] There are obviously other decision-making units in the organization who have an influence on the implementation of the application. In marketing parlance one would consider, in addition to users, the 'buyers' and 'prescribers', i.e. individuals who may have to approve the required budgets and who may give advice or specifications, respectively. They may be controllers, the users' superiors, or the personnel manager in the case of pedagogical applications. Although these 'buyers' and 'prescribers' may have an important influence on the initial implementation, the ongoing exploitation of the application and its longer-term success will depend primarily on the users.

TABLE 2. CHARACTERISTICS OF THE AGGRESSIVE AND SUPPORTIVE APPROACHES

	Aggressive approach	Supportive approach
Orientation	Task-orientation	User-orientation
Primary objective	Improvement in task performance	Satisfaction of user's needs
Potential applications	Decision-making models	Management support system
Design approach	Ideal down	Current up
Role of manager	Manager observant	Manager active
Role of analyst	Analyst active	Analyst observant
Organizational climate	Threatening	Motivating

a higher position. These needs can be satisfied effectively only if they are perceived by the manager. In order to test the importance of these needs as perceived by the manager, and not by the management scientist, it is essential that the manager estimates the resources that can be justified for the application, including the efforts and time that he is himself ready to invest in it.

Adopting broader objectives for management science implies that the scope of potential applications to be considered is also much wider than the traditional decision-making models which themselves are unfortunately often limited to stochastic, mathematical programming or simulation models depending on the prejudices of the analyst. Potential applications should, for instance, include computer-assisted learning, pedagogical games, simulation of competitive activities, automatic report generation, and a myriad of 'minor' applications which may have high practical value although they may present little technical challenge to the analysts. Together, these applications constitute a management support system.

The design of the application should start from the current situation and bring improvements within the limits of observed constraints, rather than starting from an ideal solution and compromising because of the constraints. In considering the constraints existing on the application, the analyst should give particular attention to the level of interest of the managers, their time availability and their capacity to assimilate new inputs, rather than to center uniquely on data and technique availability.

The analyst should not have a predominant role in the design and use of the application as this will most often relegate the manager to a position of observation. The manager should have an active role, and the analyst should keep a low profile. Finally, the establishment of an organizational climate where the manager is continuously motivated to exploit the benefits of the application is certainly one of the requirements of a supportive approach. The obvious danger is that the manager feels threatened by the application, as it is often the case under the aggressive approach.

The supportive approach will obviously not guarantee the success of all management science applications which, like any business venture, always comprise some elements of risk. It should, however, correct the glaring short-sightedness which has characterized management science to date, and which is similar in many points to what Levitt has called the 'marketing myopia' [17]. This approach should not only increase the likelihood of success of applications, but it should also broaden the scope of potential contributions and give a greater importance to applications which have so far remained outside the traditional stream of management science. Such an application in the area of marketing strategy formulation, MARKSTRAT, will be described in the following section.

AN EXAMPLE OF A SUPPORTIVE MANAGEMENT SCIENCE APPLICATION IN MARKETING STRATEGY FORMULATION: MARKSTRAT

MARKSTRAT is a pedagogical simulation designed to provide a training ground for the formulation and implementation of marketing

Vol. 7
1979

strategy [14]. Its development has been triggered by the following three problems often encountered by managers and educators:

1. The tendency for product and marketing line managers to neglect long-term competitive issues under the pressures of immediate operational problems.

2. The difficulty operational managers have in understanding the role of specific product-market entries within the overall corporate strategy. There appears to be a conceptual, substantive, and philosophical gap between marketing programs and corporate plans, between marketing management and corporate strategy.

3. There is also a gap between marketing strategy and marketing research. For instance, recent perceptual mapping techniques such as non-metric dimensional scaling are often exclusively used by marketing researchers for control purposes or for concept testing. They represent potentially powerful tools for the formulation of marketing strategies for multiple product-market entries, to be used by operational managers.

There are obviously several alternatives to alleviate these problems, such as the creation of procedure manuals, or the systematic use of planning models. However, it was felt that any long-term solution to these problems implies the development of the managers' capabilities, benefiting both the manager and the organization. Understanding the key issues in the formulation of marketing strategies for multiple product-market entries should indeed induce a product manager to evaluate his actions in the context of a more global corporate strategy, but should also prepare him for higher positions.

The MARKSTRAT simulation was developed in response to these business and pedagogical problems. It has been extensively used with graduate students but it was originally intended for managers in executive programs and in-house training programs, and it is this last type of application which will be discussed here.

The MARKSTRAT simulation

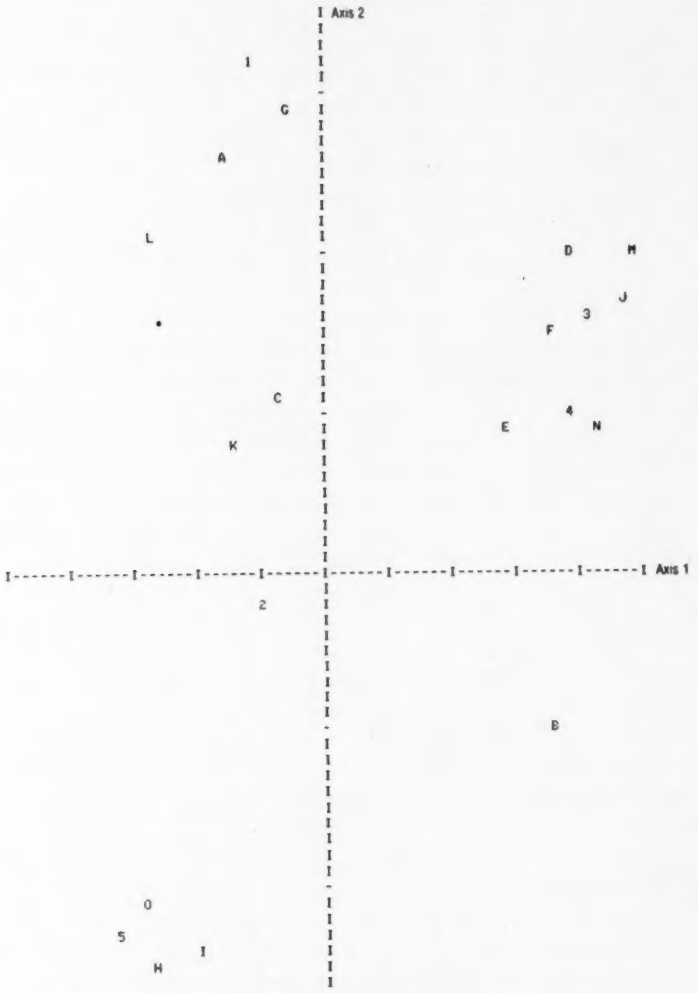
In MARKSTRAT, managers are given the management of the marketing department of a firm. There are five competing firms represented which start in different situations in terms of market shares, brand characteristics, awareness, positioning, distribution coverage, profitability and other factors. Each simulated period covers a full year of operations and the game is run over six to ten periods. In each period, the firms have to make decisions concerning brand management, the sales force, new product development and marketing research. At the end of each period, they receive a detailed company report as well as the marketing research which they have purchased and which may include awareness, shopping habits and purchase intentions surveys; consumer and distributor panels; semantic scales; perceptual mapping; market forecasts; estimates of competitive advertising and sales force; as well as advertising and sales force experiments.

The original market is composed of several segments which have different needs and behave differently. In addition, the firms may choose to enter and develop a new market for a new concept. They can ask R & D to develop new products with specific characteristics and introduce them. The firms operate in a dynamic environment with changes in productivity, inflation, changing GNP growth rates, and possible price controls. The simulated market situation will evolve according to the actions taken by the firms, and the various product-markets will consequently have different life cycles in independent administrations of the game, depending on the behavior of individual firms and of the industry as a whole.

The key issue for the MARKSTRAT firms is the selection of product-market strategies, and this is pedagogically reinforced by the availability of graphical representations of brand positions and segment ideal points through perceptual mapping, as represented by the example on Fig. 1. These strategies have to be translated into long-term plans and contingency plans. They are implemented through interaction with other departments of the firm and decisions on brands, advertising budgets and objectives, pricing, distribution, sales force and marketing research.

The impact of MARKSTRAT

Although it is always difficult to evaluate the



IDEAL POINTS	SEGMENT	COORD.AXIS1	COORD.AXIS2
1	1	-4.4	15.8
2	2	-4.2	-1.0
3	3	16.3	7.5
4	4	15.5	4.8
5	5	-13.2	-11.9

POSITIONING	BRAND	COORD.AXIS1	COORD.AXIS2	COMMENTS
A	SAMA	-6.1	12.8	
B	SALT	14.5	-5.0	
C	SALK	-3.0	5.5	
D	SARE	15.5	10.0	
E	SEMI	11.5	4.5	
F	SELF	14.2	7.5	
G	SETA	-2.2	14.4	
H	SIRO	-11.0	-12.8	
I	SIBI	-7.7	-12.5	
J	SIRU	18.6	8.1	
K	SOLD	-5.7	3.5	
L	SONO	-10.8	10.1	
M	SONY	19.7	9.5	
N	SULI	17.7	4.3	
O	SUSI	-11.5	-10.5	

FIG. 1. *Perceptual mapping in MARKSTRAT.*

Vol. 7
1979

impact of a new managerial or pedagogical tool, Table 3 provides a framework to compare the potential contribution of alternative approaches to the acquisition of marketing strategy skills. The horizontal dimension represents a crude division of the marketing process into four stages: strategic analysis, strategy formulation, programs formulation and control. In a way similar to the 'hierarchy of effects' often used in advertising [16], the vertical dimension distinguishes three stages in the adoption process: cognitive, affective, and active. Cognition is the realm of thought and the cognitive stage corresponds to the learning of the existence of new concepts and the acquisition of knowledge. The individual may then evaluate them and develop attitudes towards them, and this is the affective stage. Finally, he can go one step further in the adoption process, and use these concepts in his actions.

Obviously, the end objective of management science and, more generally, of management education is to have an impact on actions, taken by managers. Unfortunately, the only impact which is often observed is at the cognitive level, i.e. the awareness and knowledge of the concepts or the tools, not their use in actual practice.

In the context of this framework, the primary functions of three educational approaches in marketing strategy are represented in Table 3: Traditional Educational Methods, Computer-Assisted Learning and Marketing Management Games.

The positioning of each of these approaches within this 'impact matrix' is only illustrative. It only represents the primary impact as all approaches have certainly, with varying degrees, an impact at all levels. In addition, within each of these approaches, it is always possible to find specific examples which do not follow the projected typical pattern.

With these words of caution, one may consider that Traditional Educational Methods (e.g. lectures and reading) are very flexible and can cover all areas of the marketing process. Their main impact is however limited to the cognitive level. Computer-Assisted Learning (e.g. educational models) are much more limited in scope and center usually on the formulation of marketing programs. This approach allows, however, the manager to evaluate concepts and to actually use them in a simulated environment. Finally, Marketing Management Games add to educational models the possibility to have a sequence of decision periods in a dynamic environment, and thus to illustrate marketing control concepts.

Within the framework represented in Table 3, it appears that the above educational approaches do not provide an important impact beyond the cognitive level in the areas of strategic analysis and strategy formulation. Marketing Management Games provide a satisfactory environment to reach an active level of adoption, but they are usually too simple to illustrate concepts in strategic analy-

TABLE 3. FRAMEWORK FOR THE EVALUATION OF PEDAGOGICAL METHODS IN MARKETING STRATEGY

		Stages in the marketing process			
		Marketing strategic analysis	Marketing strategy formulation	Marketing programs formulation	Marketing control
Stages in Adoption process	Cognitive	TEM	TEM	TEM CAL MMG	TEM MMG
	Affective			CAL MMG	MMG
	Active			CAL MMG	MMG

TEM: Traditional Educational Methods (lectures, readings).
CAL: Computer-Assisted Learning (educational models).
MMG: Marketing Management Games (emphasis on marketing mix).

sis and strategy formulation. The important 'gap' identified in Table 3 is a pedagogical approach which addresses the Affective and Active levels of Strategic Analysis and Strategy Formulation.

This framework at the same time shows the positioning of MARKSTRAT compared to alternate approaches and also provides some further explanations for its success in practice with managers. MARKSTRAT has been designed specifically to fill the gaps described above and to provide an active level of adoption of concepts in all four stages of the marketing process. This has been achieved by incorporating in a game structure a number of elements to allow the illustration, evaluation and use of concepts in strategic analysis and strategy formulation. Some of these elements are, for instance, multiple segments with changing characteristics, the possibility of launching new products, and marketing research studies for positioning purposes.

MARKSTRAT has now been implemented in a number of executive and in-house training programmes. The reaction of the participants and sponsors has always been consistently very positive. Managers become highly interested, motivated and involved in the simulation. They acquire a deeper sensitivity to strategic marketing issues. Although the simulation represents a hypothetical durable consumer goods industry, participants actively suggest how the concepts acquired may be transposed to their own businesses which have been as varied as the food and hotel industries in past applications of the game. Finally, the most gratifying experience for management scientists and educators is to observe the impact of the application on the *behavior* of the managers, rather than on the mere acquisition of knowledge. Participation in the MARKSTRAT exercise certainly affects the behavior of managers. This is observed as the simulation progresses through successive periods and in discussions on marketing issues which follow the simulation. In some instances, managers have changed their marketing research plans, following participation in the MARKSTRAT exercise, to address issues of greater relevance for marketing strategy formulation.

The supportive nature of MARKSTRAT

It is our contention that the MARKSTRAT simulation owes much of its success and impact on managers to the supportive nature of the approach that it typifies. The objective of the simulation is not to solve a specific problem but to help managers to understand better the formulation of product-market strategies and their translation into marketing plans and programs. Companies which have sponsored applications of the simulation have done so to induce managers to think more in terms of strategic issues than what they tend to do in the operations of their business. But MARKSTRAT provides benefits to the end user, the manager participating in the exercise, not only in terms of his ability to solve current problems but also in terms of his personal competence and his professional career. In the design of the application, constraints at the user's level were explicitly considered. A game simulation was selected over an interactive model as it generates more interest, creates a more motivating environment, and is less threatening. The gaming approach was also particularly attractive as it creates a process in which new behavioral patterns will result in changes in attitudes and knowledge, contrasting with the opposite process which is often implicit in most learning situations. Particular attention was paid to the complexity of the simulation to reach a satisfactory compromise between realism and pedagogical effectiveness. Despite these *a priori* considerations, some elements of the simulation were simplified to improve the value of the exercise, after experience was gained in the first applications of the game. The simulation does not force managers to use sophisticated tools or concepts. It provides them an opportunity to do so if they wish. They will typically start to use their own implicit models and then try new approaches, test their value, and draw conclusions themselves. The simulation is disconnected, in its contents and timing, from the decision-making pressures placed on managers. Finally, the manager has the active role in the simulation and is motivated to, while the analyst observes and participates when he is called on.

CONCLUSIONS

The most striking aspect of an application such as MARKSTRAT is that it does not

Vol. 7
1979

appear to be a regular application of management science since it does not address a specific problem in need of a solution. Technically, it is based on a widely used tool, simulation, and was developed through the traditional process of analysis, formulation, programming, parameter setting, testing and documentation. It may even be considered as a major application as it required two full-time equivalent man-years and represents several thousand lines of computer code. The major difference lies, however, in the approach and not in the technique.

This approach which has been labelled 'supportive' centers on the needs and characteristics of managers as users of the application, rather than on the solution of a specific decision-making problem. It provides a framework to plan the implementation of applications in such a way as to increase the likelihood of their acceptance by managers. An example of such an approach is presented by Davis and Taylor [7] who have developed a pedagogical simulation and administered it to a group of managers before the introduction of a production line-balancing model. They have found that the game has 'unfrozen' the prevailing decision-making situation, created a higher degree of involvement of managers in the development of the model, and provided a structure to generate feedback to the management scientists.

But a more important contribution of the supportive approach in the long run may be that it defines a wide scope for management science applications. Rather than limiting these applications to traditional decision-making or problem-solving models, the supportive approach broadens the mission of management science to the development of managers' capabilities. In particular, it brings to light a largely unexploited, but substantial, educational potential for management science in business. Perhaps management scientists should reconsider their role in the light of the words expressed by Kuan-Tzu more than two thousand years ago: "If you give a man a fish, he will have a single meal. If you teach him how to fish, he will eat all his life."

REFERENCES

1. Aaker DA & Weinberg CB (1975) Interactive marketing models. *Jl Mktg* **30**(4), 16–23.
2. Amstutz AE (1967) *Computer Simulation of Competitive Market Response*. MIT Press, Cambridge, Massachusetts.
3. Assmus G (1975) The design and implementation of a new product model. *Jl Mktg* **30**(1), 16–23.
4. Boulden JB (1971) Computerized corporate planning. *Long-Range Plann.* **3**(4), 2–9.
5. Buzzel RD (1964) *Mathematical Models and Marketing Research*. Boston: Div. Res., Grad. Sch. Busin. Admin., Harvard University.
6. Claycamp HJ & Amstutz AE (1968) *Simulation techniques in the analysis of marketing strategy*, from *Applications of the Sciences in Marketing Management* (eds FM Bass *et al.*), Wiley, New York.
7. Davis KR & Taylor BW (1976) Addressing the implementation problem: a gaming approach. *Decis. Sci.* **7**(4), 677–688.
8. Edelman F (1965) Art and science of competitive bidding. *Harv. Busin. Rev.* **33**(4), 53–66.
9. Gershefski GW (1970) Corporate models: the state of the art. *Mgmt Sci.* **16**(6), 303–311.
10. Grinyer PH & Wooller J (1975) Computer models for corporate planning. *Long-Range Plann.* **18**(1), 14–25.
11. Hayes RH & Nolan RL (1974) What kind of corporate modeling functions best? *Harv. Busin. Rev.* **52**(3), 102–112.
12. Holloway C (1974) Developing planning models. *Long-Range Plann.* **7**(1), 52–57.
13. Kotler P (1971) *Marketing Decision-Making: A Model Building Approach*. Holt, Rinehart & Winston, New York.
14. Larréché JC & Gatignon H (1977) *MARKSTRAT: A Marketing Strategy Game*. The Scientific Press, Palo Alto, California.
15. Larréché JC & Montgomery DB (1977) A framework for the comparison of marketing models: a Delphi study. *Jl Mktg Res.* **14**(4), 487–498.
16. Lavidge RJ & Steiner GA (1961) A model for predictive measurements of advertising effectiveness. *Jl Mktg* **25**(6), 59–62.
17. Levitt T (1960) Marketing myopia. *Harv. Busin. Rev.* **38**(4), 45–56.
18. Lewin K (1951) *Field Theory in Social Science*. Harper, New York.
19. Little JDC (1970) Models and managers: the concept of a decision calculus. *Mgmt Sci.* **16**(8), 466–485.
20. Little JDC (1975) Brandaid: a marketing-mix model, part 1: structure. *Ops Res.* **23**(4), 628–655.
21. Little JDC (1975) Brandaid: a marketing-mix model, part 2: implementation, calibration and case study. *Ops Res.* **23**(4), 656–673.
22. Little JDC & Lodish LM (1969) A media planning calculus. *Ops Res.* **17**(1), 1–35.
23. Lodish LM (1971) CALLPLAN: an interactive salesman's call planning system. *Mgmt Sci.* 18(4), Pt 2, 25–40.
24. Meador CL & Ness ON (1974) Decision support systems: an application to corporate planning. *Sloan Mgmt Rev.* **15**(2), 51–68.
25. Montgomery DB, Silk AJ & Zaragoza CE (1971) A multiple-product sales force allocation model. *Mgmt Sci.* 18(4), Pt 2, 3–24.
26. Montgomery DB & Urban GL (1969) *Management Science in Marketing*. Prentice Hall, Englewood Cliffs, New Jersey.
27. Montgomery DB & Weinberg CB (1973) Modeling marketing phenomena: a management perspective. *Jl Contemp. Busin.* **2**(4), 17–43.
28. Naylor TH & Schauland H (1976) A survey of users

of corporate planning models. *Mgmt Sci.* **22**(9), 927–937.
29. OPPLIGER PE & WAGNER GR (1977) *Implementing a computer-assisted planning system at Dow Corning.* Paper presented at TIMS/ORSA meeting, San Francisco, May 9–11.
30. PARSONS LJ & SCHULTZ RL (1976) *Marketing Models and Econometric Research.* North Holland, New York.
31. SIMON LS, LAMAR C & HAINES GH (1976) Managers' uses of models. *Omega* **4**(3), 253–264.
32. URBAN GL (1970) Sprinter mod. III: a model for the analysis of new frequently purchased consumer products. *Ops Res.* **18**(5), 805–854.
33. URBAN GL & KARASH R (1971) Evolutionary model building. *Jl Mktg Res.* **8**(1), 62–64.
34. WAGLE B (1969) The use of models for environmental forecasting and corporate planning. *Opl Res. Q.* **20**(3), 327–336.

ADDRESS FOR CORRESPONDENCE: *Professor Jean-Claude Larréché, Institut Européen d'Administration des Affaires, Boulevard de Constance, F-77305, Fontainebleau Cedex, France.*

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 3. pp. 219 to 231
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0601-0219$02.00/0

The Status of the Systems Approach

RA BRYER

University of Warwick

(*Received July* 1978: *in revised form December* 1978)

This paper is an assessment of the attempts by Ackoff, Beer and Churchman to provide a justification for the claim that OR is a science. These writers tackle the problem using the philosophies of positivism, pragmatism and conventionalism/idealism in an attempt to build an ideal scientific methodology for OR. The conclusions reached are that no such methodology can exist and that philosophy confuses rather than clarifies the nature of OR as a science. The weakness of the philosophical approach to some of the problems faced by OR is demonstrated through a discussion of Ackoff and Emery's analysis 'On Purposeful Systems'. An outline is given of an alternative framework through which the problems of practising OR can more fruitfully be viewed.

ol. 7
979

INTRODUCTION

IN A RECENT work Ackoff [7] appears to argue that there are no natural limits to the potential scope and usefulness of OR-type activities. In his and others' eyes it can justifiably claim the competence to play a significant, if not always determining, role in all public and private policy issues, ranging from the organisation and control of our educational, criminal, transportation, urban, waste-disposal and other such systems, to the more usual concerns with the whole range of corporate strategies and tactics [*cf.* 4].

In the OR community at large far more effort has been devoted to promoting its aspirations and past achievements than with providing solid reasons why any client should feel compelled to accept its findings and recommendations. Typically this issue has been sidestepped by arguing either that OR only proposes whereas it is the role of the client (as decision-maker) to dispose; or that resistance is merely a problem of 'implementation', which requires a secondary level of 'educational' activities directed at the client. In recognition of the inadequacy of this type of response Churchman and Schainblatt [20] have proposed an extension of 'education' activities to include the understanding of the client by the operational researcher. Only when researcher and client reach a 'mutual understanding', Churchman and Schainblatt argue, should the client feel compelled to accept the solution which the researcher also should, under those conditions, feel compelled to propose. However, this solution to the client's problem leaves open the question of *how* he can judge whether the researcher 'really' understands him. That is, in what sense *should* the client understand the operations researcher?

Before the client should feel any compulsion to accept the researcher's conclusions he has the inalienable right to receive assurance not only about the researcher's good intentions towards him but also about the *quality* of his understanding. It is here that the often made claim that OR is 'scientific' plays a potentially important role.

To this point in its history the OR community has traded heavily upon and actively promoted a belief in its scientificity. It has become a core element in its self-conception [*cf.* 12] and the pressure which a scientific aura exerts on the commissioning and acceptance of research outcomes stands out as a possible topic for future empirical work [*cf.* 24]. It is only on the basis of this and other related work that OR could ever fully understand its actual rather than its pre-conceived role in the improvement of social systems [*cf.* 25]. The appreciation which its clients have of 'science'

may, in practice, influence the type of problems which OR is asked to work upon and also the acceptability of the results.

However, some of the OR community have seen this type of work as only *one* of the 'researchable' aspects of the implementation problem. It has appeared to them much more important that OR provides its clients with a fully reasoned justification of its scientificity. The aim of this work has been to provide such a convincing account that the client would feel *compelled* to accept results that OR, as such a 'science', produced.

Considering its apparent potential for the development of OR only a very small number of its community has even attempted this type of research, and the work of those few who have has gone largely undiscussed. Notable contributors to it are Ackoff, Beer and Churchman, and this paper is critical assessment of some of the key elements of their work.

Although these writers proceed from a variety of different directions they all reach the same conclusion, *viz*, that the *only* way in which OR's scientificity could ever be justified is in terms of the *methods* which it uses for research, i.e. its 'scientific methodology'. The implicit line of argument here is that if the client accepted and 'believed in' the methodology he should also accept and believe that the results produced by it are useful to him—on the argument that only if a result was useful to him would the client feel compelled to accept it. The viability of this approach clearly depends upon the demonstration of a *logical* connection between the method of inquiry and the usefulness of the results. Thus a failure in *logical* coherence vitiates the whole project. The long-standing belief of most philosophers of science that there was such a logic of science which underlay, as their essence, the practice of the individual sciences, is now seriously challenged (e.g. [23, 29]). And yet *this* is the approach favoured by writers on the nature of OR.

It would be interesting to know what impact these writers had on the intellectual frameworks which guide practising operations researchers in their day-to-day work. Clearly, there are many other possible sources upon which they could draw (not least of which are the 'cultures' of the organisations within which they work), and an understanding of them would allow OR to claim that it really pursues its inquiries self-consciously. However, it is the potential impact of academic ideas on practical men which is of interest here, and who else but to Ackoff, Beer and Churchman can they turn for practical guidance? These writers have self-consciously turned to philosophy to provide this. Operational researchers are practical men seeking sound practical advice, and they cannot realistically be expected to check for themselves whether or not by accepting it they are unwittingly becoming (to rephrase Keynes' words) "... the slaves of some defunct philosopher" [27]. In the hands of Ackoff, Beer and Churchman they are, and a major function of this paper is to make them aware of the fact.

In the final section an alternative to the philosophical approach is outlined. It is based upon the belief that the best way to a client's heart is the *direct* one. Clients, as purposeful systems, need 'appropriate' knowledge to enable them to pursue 'appropriate' goals. Therefore, attempts to systematise the practice of OR could usefully focus directly on the problems of clarifying and extending the objectives which clients should and do pursue and the types of knowledge which are appropriate to these.

The *indirect* approach to the client favoured by OR's theoreticians can be seen to draw upon the established, but now largely defunct, philosophies of science-positivism, pragmatism and conventionalism/idealism. A key feature of all of these philosophies is their degradation of the accumulated knowledge of the established sciences in their vision of the essence of science. This is surprising because the success of any science can be reasonably, if loosely, defined as the adequacy of its accumulated knowledge for the objectives which it pursues [*cf.* 29]. No one should expect OR to pursue unquestioningly the objectives which others have set for themselves, but they might reasonably ask OR's theoreticians to find general ways of adapting them and the associated knowledge to their client's objectives. So far no systematic attempt has been made to do this, and therefore we are forced to judge OR's theory solely in terms of the logic of its proposed methodology. The conclusion will be that this is as unsatisfactory as the philosophies of science upon which it is based.

Before we turn to analyse these philosophic-

Vol. 7
1979

two further comments must be made. Firstly, as the discussion progresses it will become clear that most if not all of the blame for the malaise of OR's theory must be laid at the door of the systems approach. Secondly, there is little hope that developments of the systems approach could remedy this situation [*cf.* 37]. This is so because it is the systems approach itself which engenders the philosophical outlook: the troubles which it creates for its adherents are internally generated. In a previous paper [13] the viability of the systems approach as a basis for the successful practice of OR was shown to depend upon its ability to incorporate within its framework such obviously important real-world phenomena as the generation and exercise of power in social organisations. It was suggested there that this seemed unlikely. In this paper theoretical analysis of the systems approach and its application to social organisations (in the hands of Ackoff and Emery [10]) justifies that earlier suggestion. This analysis therefore starts with an examination of the uses to which the systems approach has been put and the problems which it creates.

USES OF THE SYSTEMS APPROACH

Ackoff, Beer and Churchman appear to argue that OR's rapid development depends upon gaining widespread agreement in the OR community about the objective which it should pursue and the general means which should be employed. Much of their work is intended to persuade the OR community to accept their vision of these.

They have found the 'systems approach' very much more useful in the specification and justification of an objective for OR than it has been for the generation and justification of appropriate means. In other words, the systems approach is very likely to convince the client that his best interests are being pursued, but not that they would be met. Although it has provided an intellectually unassailable objective for OR, in doing this it has simultaneously made it practically unobtainable. This result, now widely recognised in the OR community at least, is the reason why its objective is currently put forward as an *ideal*, i.e. as a highly valued objective which it is impossible to achieve in practice. The implication of this result is not, in itself, grounds for abandoning the ideal, but because means do not flow naturally from it, extra care must be exercised in their choice so that they are at least *consistent* with it. So far no one has found a way of doing this nor, we shall see, are they ever likely to.

This is because, fundamentally, the systems approach amounts to the injunction to deal with *nothing less* than 'complex-totalities' in any attempt at system improvement. This view is now widely accepted in the OR community. Ackoff and Rivett [9], for example, argue that OR adopts the "... system's orientation" when "... it deliberately expands and complicates the statements of problems until all the significantly interacting components are contained within it" [9, p. 10]. The ideal of OR is to grasp the 'whole system' because the justification of a claim to have really 'improved' a particular situation is dependent upon an understanding of how seemingly disparate 'parts' are related as a 'totality'. Nothing less will satisfy the systems approach.

Even the seemingly practical rule of thumb which suggests that problem definitions should be expanded until no significant difference to the problem solution results [e.g. 11] cannot escape the whole system ideal. This suggestion is only of practical use *after* the whole problem has been defined because without knowledge of the whole system we could not be sure that an insignificant result was foreshadowing a significant one [*cf.* 19]. Churchman has gone so far as to argue that dealing with this type of 'wicked problem' is a *moral* imperative for OR [16], if it accepts, as he argues it must, a "... monistic philosophy: there is one world of interconnected entities, not many" [17]. For Churchman "... the joy of OR is that it is in the centre of the deepest mysteries of the human race, because, academically speaking, it has taken on the whole system" [18, pp. 8–53].

No serious challenge has been or is likely to be mounted to the whole-system ideal in its own terms because, in essence, it is a tautology. To exclude potentially significant knowledge from a problem definition on the grounds that it is *in principle* irrelevant to real improvement is obvious logical nonsense.

The significance of the so-called 'alternative' approaches of satisficing and incrementalism is merely their different view of what is practically attainable, and not their view of what is

desirable. However, although these alternative approaches do not challenge the whole system ideal they do point to its weakness as a practical guide for action. How are we to approach the whole system ideal when seemingly *any* attempt to lay down guidelines to the investigation of systems must appear, in its terms, to be an *arbitrary* exclusion of potentially relevant knowledge? If we take the whole system ideal seriously it appears that we must have full knowledge before we can provide guidelines for inquiry. However, if we had this knowledge inquiry would be unnecessary!

The tautological character of the whole system ideal has provided very solid ground from which to attack a variety of attempts to limit the freedom of OR to inquiry into the full complexity of problem situations. In particular it prohibits operational researchers accepting problems whose nature is pre-defined for them. Because Ackoff [1] accepts the whole system ideal that "... improv(ing) an organisation's performance ... can only be effectively solved by an integrated examination of the organisation" [1, p. 17] he can easily conclude that "where ... (problem) ... reductions are imposed on the researcher by external forces no problem arises other than the one of determining whether the problem should be tackled under the imposed constraints" [1, p. 30]. Beer adopts the same view: "the point to be made is this: operational research must encompass the *whole* of the problem situation, and management may not succeed in defining what this is" [11, p. 50].

The application of given techniques for problem solution gets the same short shrift, for the same reason, from these and other writers e.g. [35, 36]. The significance of these attacks is that they are symptomatic of the irresistible pressure of the systems approach against any attempt to 'understand' a problem situation before it is thoroughly studied. The lesson is a salutory one but the unremitting emphasis which these writers place on it makes very urgent the question of what guidelines *can* operational researchers follow in their inquiries without thereby subverting their ideal?

The prominent writers on OR have all concluded that the solution to this problem is a *method* of inquiry which does not depend upon a view of the whole system. This has been interpreted to mean a method which is justified *in its own right*, regardless of the subject-matter to which it is applied. However, the philosophies of science which are turned to in this quest are made of the same stuff as the systems approach. They, too, demand that knowledge shall be *complete* before it is pronounced true. To compromise the systems approach is to abandon it; so too with philosophies of science. Once one steps into the circle of this logic there is no way out, but there is also no way forward. Astute clients will perceive this and remain agnostic.

OR AND THE PHILOSOPHY OF SCIENCE

Writers on the methodology of OR have, not surprisingly, drawn their inspiration from the major philosophical options of today, namely, positivism, pragmatism and conventionalism/idealism. These philosophies offer methodological rules which they claim can be applied to determine the validity of any knowledge. These rules do not form mutually exclusive sets and within broad limits it is possible to consistently select elements from each. Writers on OR invariably do this.

Positivism prescribes that the limit of valid knowledge is 'phenomena', i.e. that which is actually experienced. Positivists claim that any attempt to (as it were) 'go beyond' the phenomena of experience immediately makes knowledge metaphysical and non-scientific. The import of this constraint is that we are forbidden to ask *what* it is that produced the data that we observe. This type of question is ruled out as 'meaningless'. The status which 'theories' have, the positivist claims, is limited to their role as ways of organising experience. Theories do not, in other words, describe any reality beyond the experiences which they represent. A corollary of these two injunctions is that what is really significant about scientific inquiry is only *method*; questions about the 'nature' of phenomena or the 'nature' of theories are considered meaningless by positivists. In other words, it is an implication of positivism that an assessment of the scientificity of a discipline of inquiry is only to be made by whether or not it employs an undifferentiated and unifying method, and not by the body of knowledge which it produces. For positivists the 'whole system' is merely all

Vol. 7
1979

aspects of all phenomena and so the scientificity of a discipline of inquiry is not determined by which phenomena (or aspects of phenomena) are chosen for study. For positivists, therefore, the problem of grasping the whole system is merely a problem of 'coverage'.

Positivism is clearly a contender as a method of inquiry with which OR could pursue the whole system ideal because no prior commitment to a view of the whole system appears to be required. Its influence can in fact be detected. A major example of this is Ackoff's emphasis on his belief that "the division of science into disciplines was accomplished by man, not by nature", even though he had previously argued in the same paper that scientific status depends in part on a discipline possessing a distinctive subject-matter. He goes so far as to argue that "the disciplines cannot be individuated by a unique set of objects or events which constitute their subject-matter, but they are distinguished by the aspect of phenomena to which they direct their attention" [1, p. 24]. It is on this *philosophical* basis that Ackoff justifies the idea of OR as an 'inter-discipline': the complex totality ideal can be met by the employment of an inter-disciplinary team because, as Ackoff says, "in the operation of a system... the aspects... (of phenomena)... of the system which can be manipulated so as to improve its performance are likely to come from many different disciplines" [1, p. 25]. Similarly, Beer argues that "... nature is the subject-matter of science" [11, p. 55], and that the grouping of scientific knowledge around subjects is merely a historical accident which has meant an unfortunate "... hardening of the faculties". Only for the technologist is it relevant to ask "'What is your subject?'" because the real scientist has to reply "'I investigate the world'" [11, p. 5]. For Beer the raison d'être of an inter-disciplinary team appears to be that each member will neutralise the 'bias' of the others [11, p. 50].

The message seems to be that OR is not interested in scientific *knowledge* as such. This is clearly spelled out by Beer when he says of the OR scientist that

> "He has no interest in arguing about the 'right' use of... (for example)... a psychological term, but only in the results. So the OR man is a special kind of scientist, for he does not have to bother with determining the laws governing basic natural phenomena... (the OR man)... must operate *across* the various scientific disciplines..." [11, p. 119].

Similarly, Ackoff argues that OR is "... not... concerned... with the body of information and knowledge which... (science) ... has generated; that is, not with specific theories, laws and facts that have been developed in the various physical, life, and behavioural sciences". "Instead", he argues, OR is "... concerned with the *procedures* by which science generates this body of knowledge, the *process* of inquiry" [2, p 5] (my emphasis).

Both the strong emphasis on the role of theory ('academic disciplines') as a mere organiser of phenomena and the striking degradation of the role of a knowledgeable grasp of particular subject-matters in the practice of science betrays a clear, but not exclusive, (see below), positivist influence on these writers.

However, any attempt to base OR's systems approach on positivist philosophy can be subjected to fatal criticisms (as Churchman, for one, has pointed out [18]). Firstly, no 'meta' logic of science has ever been satisfactorily developed, and so if this tack is chosen the 'systems approach' amounts only to an as yet unfulfilled promise, (more on this later). Secondly, there are serious doubts about even the possibility of developing such a logic.

The basis for this doubt is that such a logic would have to rest on what has come to be called a 'neutral observational language'. Positivists claim that facts and theories constitute qualitatively *independent* domains and, as the logic of science is the logic of theory choice, some means of transcending these domains is therefore required. This is the envisaged role for a neutral observational language. However, it has been pointed out by some philosophers that there is no such thing as 'pure experience'—theory-free observations. Because of the inextricable relationship between facts and theories they cannot be judged by the facts alone, because the attempt to do so involves a circular dependence upon either the theory in question or other theories. Thus, there could not possibly be a positivist systems approach because "... the data of experience always leave scope for more than one explanatory hypothesis, and which one is chosen cannot be determined by experience" [28, p. 159]. Because theories cannot be tested by the facts

alone, the attempt to do so means that the larger view (the whole system) within which they make 'sense' is taken for granted, a point which Churchman has often made. This criticism is fatal because for OR the larger system is, according to the systems approach, the object of inquiry.

For positivists the larger system does not exist as such; it exists only as 'phenomena'. For them there are individual 'sciences' only in the sense that groups of people focus their attention on particular clusters of phenomena. As far as the positivist is concerned, because the theories of the specialist sciences merely 'mediate' their particular phenomena, they are not themselves a part of 'knowledge'. As Kolakowski points out, for the positivist the theoretical knowledge produced by the various sciences is merely "... a collection of guidelines, useful and indispensable in practice, but devoid of truth value" [28, p. 52]. And so it is for Ackoff and Beer, as we have seen. Their phenomenalist emphasis is, then, very definitely positivist. However, their emphasis on method is not because on this topic positivism has little to offer. Their elaboration of a methodology for OR has, in fact, drawn upon pragmatism.

At first glance pragmatism and positivism appear to be completely distinct philosophies, in the sense at least that pragmatism argues that the real meaning of theories is their *practical application*, and not merely their organisation of sense data. However, Kolakowski [28], amongst others [e.g. 23], has argued that at a deeper level pragmatism is merely the translation of positivism into a 'way of life'. Both Ackoff and Beer are self-declared pragmatists, and Beer explicitly bases his argument for the scientificity of OR on the founder of pragmatism, Charles Peirce [see: 11]. An analysis of Peirce's work shows quite clearly its positivist foundations. (My analysis closely follows that of Habermas [23]).

On the face of it pragmatism is ideally suited to meet the needs which OR's theorists have seen the systems approach create. This is so because it directs attention away from the positivistic focus upon the logical structure of theories towards a concern with the methods by which theories are generated and tested. The appeal of Peirce's work to OR is also strengthened because of his commitment to the view which sees science as an organic, adaptive process which inexorably increases man's power to control the world. The basic pragmatist assumption (as Habermas notes) is "that we may rely on the effective functioning of ... (science as) ... a self-regulating, cummulative learning process" [23, p. 93].

From this basic value-judgement is derived the role of the philosophy of science. Its goal is the demonstration of the truth of the assumption. Peirce and his followers attempt to achieve this through an analysis of the logic of the methods of science. Thus, pragmatists claim that there is a "logic of scientific progress" [23, p. 94].

The specific philosophical task which pragmatists face is a demonstration that 'methodology' can produce 'reality'. This is, to say the least, very difficult, and it is made even harder for Peirce because of his determination to distinguish his work from positivism. To do this he heroically abandons the notion of reality as existing 'in itself' and resolves to talk only about methodology (*cf.* Churchman [15]—who resolves that 'facts', from a methodological viewpoint, are 'theories'). However, remaining entirely faithful to methodology creates insoluble problems. The most important of these is that without some knowledge of reality that is independent of our methodology (which Peirce describes at great length, but whose specifics are not directly relevant here) we cannot know if our methodology puts us in touch with 'reality'.

The pragmatists answer to this is a reassertion of their basic assumption: science just *is* progressive in the long-run, and at the end of the process of inquiry as a whole we will be forced to conclude that we have grasped reality. Therefore, they argue, reality is, in the long-run, a 'correlate' of methodology. However, the 'long-run' is a 'long-way' off, and we clearly need some assurance that usable knowledge will be produced in the interim. Therefore, Habermas notes that pragmatists must also "... assume at the same time that, in increasing measure, we objectively arrive at true statements even *before* the consummation of this process—despite subjective uncertainty about the truth value of every single one of these statements" [23, p. 110]. Churchman in particular has come across this difficulty in his own work, and has been unable to find a satisfactory resolution of it (see below).

Vol. 7
1979

In the face of this problem it is not logical to argue—as both Ackoff and Beer appear to do—that, nonetheless, the methods of science do produce an understanding of reality because, referring to Habermas again, "if we assume that reality is not constituted independently of the rules to which the process of inquiry is subject, then we cannot refer to this reality to justify the validity of the rules of the process..." [23, p. 118].

So, the theories of measurement put forward by both Ackoff [2] and Churchman [15] on the pragmatist 'ticket' stand in need of more justification than they themselves provide. Churchman in particular [17, 19] has come to realise that the validity of measurement is conditional upon a valid conception of the whole system. This has led him to promote Hegelian philosophy [19] as a possible solution because, as Habermas points out, "...the identity of concept and object, which Peirce had first derived from a methodological conception of truth and thus understood as an *interpretation of the fact of scientific progress*, can only be justified in terms of an idealism that is not unlike Hegel's" [23, p. 111]. There are, however, great difficulties involved in taking this route, not least of which is that "...the conception of a progressive embodiment of ideas, which dominates Peirce's later philosophy...(and Churchman's work)..., cannot evade the difficult concept of nature as an absolute-subject" [23, p. 111], (my insertion). Only Churchman has faced up to this necessity, but he has been unable to show how the use of Hegel's philosophy can avoid the presupposition of the whole system and so provide a methodology for investigating it.

Even though the logic of his argument led him there, Peirce did not himself pursue the line which Churchman has picked up. Instead he gives up his attempt to justify the objectivity of inquiry in terms of its intrinsic logic, and he turns instead to the 'logic of language'. He claims, but is unable to show [23, p. 107] that language is gifted with an 'index property' which means that it 'indirectly' captures the "facticity of reality". Ackoff [particularly 10] and Beer [11] both rely on this kind of argument. Beer explicitly follows [11, p. 121] Peirce and Ackoff vigorously promotes the operationalisation [10] of definitions—to what effect we shall see in the next section.

This notion of the nature of language makes it very difficult to comprehend the pragmatist programme because although it provides a guarantee of 'objectivity' to inquiry it turns the knowing mind "...into a series of empirical events,...(and)...then what remains is nothing but universal matters of fact existing in themselves and the combinations of signs through which these matters of fact are represented" [23, p. 137]. A sceptical client could be forgiven for with-holding assent to research with this methodological justification.

This particular weakness of pragmatism has prompted its supplementation by a conventionalist/idealist overlay. It finds expression in the work of virtually all serious thinkers on OR and is strongly influential in Churchman's work in particular. For the conventionalist/idealist there is a dominant role for ideas in grasping the whole system. This is because ideas are seen to determine facts (rather than *being* facts or being *determined* by them). Hence the emphasis on the importance of conceptualisation that we often find in the writings on OR.

Conventionalist/idealists charge positivists with taking the whole system for granted. A more cogent way of making this point is to notice that because positivists restrict knowledge to phenomena and treat theories as merely organising they are unable to distinguish between the accidental pairing of events and the existence of true casual relationships. In other words, positivists take the whole system for granted because they are unable and are forbidden to inquire into how phenomenal events are related. As Keat and Urry [26] put it: "...in explaining any particular phenomenon, we must not only make reference to those events which initiate the process of change: we must also give a description of that process itself. To do this, we need knowledge of the underlying mechanisms and structures that are present, and the manner in which they generate or produce the phenomenon we are trying to explain" [p. 30].

For both the positivist and the pragmatist there are no underlying realities which play a direct part in the practice of science. As regards pragmatists Kolakowski points out that for them "it makes no sense to ask 'How are things really constituted?' but only 'What do I get if I believe this or that?'" [28, p. 191]. However

the conventionalist/idealist also does not see any role for underlying realities. For example, Rivett argues that a major difficulty of doing OR is that "there is, of course, no such thing as an Operational Research problem. There are not even such things as problems, in a concrete form. The problems only exist in our minds. Hence there is a purely subjective basis to our subject" [36, p. 243]. However, placing so much stress on the subjectivity of OR raises doubts about the 'realism' of its idealisations, i.e. whether they are capable of grasping the totality and complexity of the whole system. The doubt is justified when we look at the notion of totality which has currency in OR circles.

Although we have seen Ackoff and Beer argue in positivist mode, they also sometimes argue in conventionalist/idealist mode. Thus, Beer argues that "... coherence, and pattern and purpose are, all three, acts of mental recognition rather than characteristics of physical things ... (because) ... the detection of a system in the world outside ourselves is a subjective matter" [11, p. 243]. Similarly, Ackoff argues that systems are "... *subjective* insofar as the particular configuration of elements that form ... (it and its environment) ... it dictated by the interests of the researcher" [5, p. 84]. This approach is justified only if we accept that the reality of a system can *only* be captured in its *overall* 'total' properties. That is to say, if we accept Ackoff's view that "a system is more than the sum of its parts" [6, p. 13], and that wholeness is an *emergent* property which is *primary* to the parts. This definition of a system, following as it does from a conventionalist/idealist viewpoint, denies the importance of any reality underlying the whole, *as this is conceived by the researcher.* In particular, it denies the reality of the system's 'structure' which would seem to be the very thing which we require to understand, if we are to understand it as a complex totality. Compare Ackoff's view of totality with the "operational structuralism" described by Piaget:

> "Over and beyond the schemes of atomist association on the one hand and emergent totalities on the other, there is, however, a third, that of operational structuralism. It adopts from the start a relational perspective, according to which it is neither the elements nor a whole that comes about in a manner one knows not how, but the relations among elements that count. In other words the logical procedures or natural processes by which the whole is formed are primary, not the whole, which is consequent on the system's laws of composition, or the elements" [34, pp. 8–9].

With the structuralist concept of totalities we are at least led to ask how totalities (for example, organisations and their environments) are structured and how and why those structures change. In short, with such a notion of totality and complexity we are encouraged to apply and generate relevant knowledge (scientific and lay) of the systems laws of composition. With the conventionalist approach, however, we are restricted to the level of subjective ideas about 'the system'. Because they cannot obtain a warrant of (at least) presumptive objectivity by reference to real-world structures, conventionalists turn to the *way* in which the ideas are developed, and they hope to find objectivity in *this* process. Churchman's Hegelian Inquirer [19], Ackoff's Circular Organisation [6] and Checkland's version of the systems approach [4] all seem to be inspired by this hope. However, these writers have yet to provide any good reasons (either argumentative or evidential) why this hope is a realistic one.

Doubts arise because it is hard to see how these schemes avoid *any* reliance upon an implicit grasp of the nature of the system which the process of inquiry is intended to reveal. The burden of proof here lies with the proponents of such schemes. We have seen that Churchman has been unable to discharge this burden because he is forced on to the Hegelian presupposition of nature as 'absolute subject'. Unless we could convince the client to *believe* in this notion, inquiry would, in Churchman's scheme, be meaningless to him.

Ackoff's circular organisation, which is related to his earlier scheme of adaptavising [4], is also unlikely to be able to avoid making at least some assumptions about the nature of the social world. Objective problem-solving for him is, in fact, a 'correlate' of a particular form of *social organisation.* However, because his conventionalist approach forbids any direct reliance upon *knowledge* of the real structure of social systems his assumptions could only be realistic by chance.

In recent work with Emery [10] Ackoff has, in fact, applied all of his philosophical insights to the development of concepts by means of which systems scientists can analyse social systems. In this work, therefore, we have a unique

Vol. 7
1979

opportunity to examine the fruitfulness of the philosophical approach to one specific problem area—understanding social systems. Of special interest is Ackoff and Emery's determination to avoid any systematic reference to the established knowledge of the social sciences, and to build *ab initio* the 'special' concepts which they believe the systems sciences need. The motive for this is the philosophical desire to avoid taking any part of the whole system for granted. To meet this need they rely heavily upon operationalising *a priori* definitions for use in measurement. However, we shall see that, contrary to their intentions, they are forced to take their 'data' for those measurements as 'given' to them by an underlying social system. Their measurement concepts will be shown to depend upon an implicit grasp of the nature of social systems which is strongly at variance with the accumulated knowledge of the social sciences.

ACKOFF AND EMERY'S TREATMENT OF SOCIAL REALITY:

The core of Ackoff and Emery's work is the attempt to provide a system of concepts which can be used to "... open to public examination the study of the mind's inner workings", and to do this so that all reference to internal subjective states is avoided. They wish to restrict attention to observable behaviour and rigorously defined concepts by means of which the observations are ordered and manipulated (they call these concepts idealised operational definitions and measures). The major concepts which they devlop to both measure and explain human behaviour are: (a) the concept of 'familiarity' which refers to a subject's preference for certain means in pursuing a course of action, (b) the concept of 'knowledge' which refers to the subject's knowledge of the efficiencies of courses of action for certain ends, and (c) the concept of 'intention' which refers to the preference the subject has for various ends. These three concepts define the 'personality' of the subject, and they claim that it is "... an observable function that describes how an individual or system converts a choice situation into an expected relative value for himself or itself". The claim is that we can measure each 'parameter' of the 'purposive state' independently. Thus:

> "in measuring familiarity, knowledge and intention are held constant. In measuring knowledge, familiarity and intention are held constant. In measuring intention, familiarity and knowledge are held constant" [10, p. 42].

Over a large number of trials they claim (for example) that we can arrive at a relative frequency score which is a real measure of preference for ends. Can we, in fact, by restricting ourselves to observable behaviour and their concepts, derive such a measure?

Consider the problem of actually using their scheme. Suppose we wish, for example, to measure a subject's 'belief in efficiency". We are told that:

> "As long as there is the possibility that the individual in this environment is pursuing many different ends, we cannot use his behaviour directly as evidence of what efficiency he believes a course of action to have with respect to any one outcome, for we do not know with respect to which outcome his behaviour can be taken as an indicator of such belief".
> "(Thus)... to determine an individual's belief in the efficiency of a course of action for any outcome, it is necessary for us to isolate the outcome so that his choices cannot be taken to be serving any other objectives" [10, p. 91].

How would one know, in the first place, whether there was any possibility that the individual had different ends? According to their scheme the answer is simple: all we have to do is to control for beliefs in efficiency and knowledge and then measure the ends that the individual is actually pursuing in this environment. But it was beliefs in efficiences that we were concerned with in the first place! To hold them constant we have to know them but to know them we have to hold them constant! It seems as though Ackoff and Emery's scheme of measurement presupposes that the very knowledge sought for is already at hand. Exactly the same circularity is involved if we desire to measure other elements of an individual's personality.

Given that it is impossible (in particular) to measure an individual's ends, it is difficult to see how the role which Ackoff lays down for manager and management scientist of reconciling the conflict between the goals of individuals, organisations, and environment (e.g. [8]) can be met, unless it is assumed that they are in some way already systematically related. Ackoff and Emery do presume, for example, that there is a systematic dependence between

the goals of an organisation and the goals of individuals when they define an organisation as a social system with collective or common objectives. Without attempting to provide any justification for this view they assert that it is "the intention to coproduce common objectives... (which)... produces the interactions that lead individuals to cohere as a social group" [10, p. 213]. This is a very convenient assumption in view of the weakness of their measurement methodology because it is tantamount to assuming that an individual's purposes are *determined* by the collective goals, that is, the goals of the organisation. If this were true it would provide the needed fixed point for their scheme to be operative, but there can be no purely *methodological* warrant for such an assumption. Also, the previously common practice amongst some social scientists of assuming a consensus of goals and values among the members of a social system has attracted so much criticism (e.g. [21, 22, 30, 35]) that it has all but been abandoned by them. A good deal of this criticism has been focussed around the neglect of the social realities of power and conflict which it involves. The issue here is not that characterising social systems in terms of power and conflict is *necessarily* more realistic, it is that neglect of these phenomena betrays a lack of understanding of the *nature* of social reality, and hence prevents an adequate grasp of the 'whole system'.

Assuming that there is a systematic dependence between the goals of the social system and the individual members is equivalent to *assuming* that there is no real conflict between them. As Ackoff himself says, any conflict must be 'apparent' [7]. In Ackoff and Emery's conceptual universe there cannot be conflict over the "common objective" but only "... with respect to other objectives or... with respect to the courses of action by which the common objective should be pursued" [10, p. 214]. Note also that because of Ackoff's definition of a totality as an emergent property which is prior to the parts, he could not (and does not,[7]) accept the possibility of an unresolvable conflict between the parts and the whole. If he did the 'whole' would cease to exist in the sense in which he claims that it does.

His treatment of conflict in social systems again falls foul of the vicious circularity of reasoning which seems inevitable if serious reference is not made to the nature of social reality. Ackoff charges operations researchers with the task of contriving to remove any conflicts which are reconcilable 'in principle' [7]. The 'principle' he has in mind is that operations researchers should seek to discover whether a "... conflict that appears to be unresolvable at one level of means or ends is subject to resolution at a higher level of desirability" [7, p. 96]. However, as he defines the elements of a social organisation as 'subsystems' of larger 'systems' it is unlikely any conflict could be found which he would consider unresolvable.

It is an axiom of Ackoff's systems view that a 'higher' system can always be found as the only limits to systems boundaries are the subjective conceptualisations of the analyst [5]. To admit that a conflict was unresolvable in principle would mean, therefore, that his systems approach had failed. Ackoff would, no doubt, admit this as a possibility but it is entirely unhelpful to define 'real' conflict as a by-product of the *outcome* of the work of systems scientists. Ackoff provides them with a powerful motivation to tackle some forms of conflict in social organisations, but does not and cannot provide them with the tools with which to proceed.

This very one-sided treatment of conflict is matched by the total and explicit neglect of the problem of power. Ackoff and Emery "note that co-operation and conflict exhaust the ways in which one individual can affect the expected relative values of others" [10, p. 197]. However, there is an implicit definition of power in their scheme. Because "one system can affect the behaviour of another... by producing a change in one or more of the four components of the other system's choice situation, or in one or more of the other's parameters of choice" [10, p. 142], the extent to which one entity can manipulate these defines his power in the terms of their scheme. This power would, however, only be measurable in their scheme from observations of the *behaviour* of both entities. Lukes [32] amongst others [e.g. 33] has criticised this procedure on the grounds that "... the bias of the system is not sustained simply by a series of individually chosen acts, but also, most importantly, by the socially structured and culturally patterned behaviour of groups, and practices of institutions..." [32, p. 21]. In Ackoff

Vol. 7
1979

and Emery's scheme we are forced to assume that the lack of power related behaviour (particularly conflict) means that it does not exist. With their view of power we are forced into the position of saying that the only social dimension to power is its manifestation in individual acts. If, as seems likely, power operates in more subtle and covert ways by being, for example, implicit in forms of social structure, ignoring this possibility implies that in our studies we merely reproduce whatever bias there is in the system.

Recognising a distinctly social dimension to power seems indispensible for OR [13]. Unless we recognise it as a possibility we are forced to conclude that, for example, the subjective expressions of interests of individuals and groups coincides with what could be termed their 'objective' interests, that is, the goals that the individual *would have* if, for example, he had 'perfect knowledge'. Churchman has, so far, struggled unsuccessfully with this idea. For Ackoff to convince us that he had, in any particular situation, really reconciled the goals of individuals, organisations and their environments would, therefore, require a demonstration that this was achieved in terms of objective interests and not merely subjective interests. As Churchman has pointed out, it is the responsibility of the operations researcher "... not merely to translate what the client *says* his needs are, but to study these needs in relation to the broader aims of the client" [17. p. 13]. Ackoff and Emery's conceptual scheme does not allow us to investigate these 'broader aims' in terms of the individual's social situation.

To summarise and conclude this section, Ackoff and Emery's strictly philosophical approach to investigating social organisations is incapable of dealing with obviously important dimensions of social reality. This failure arises because of the neglect of the social sciences which philosophy encourages and legitimates. These authors could not reasonably expect a client to accept the results of research into social organisations on the basis of their methodological approach.

AN ALTERNATIVE APPROACH

The failure of philosophical approaches to the problem of providing clear and practical guidance to operational researchers raises the urgent question of what other general options are available and likely to be viable. It is only a partial solution to suggest that OR takes more seriously accumulated scientific knowledge because unlike most normal scientific disciplines OR's approach to the world is normative, not merely descriptive. The lesson which should be learned from the failure of these philosophies is, rather that no completely *general* grounds for establishing the truth of knowledge exists [*cf.* 31]. If OR is to justify its research it must take full cognizance of the *diversity* of grounds which exist, and work out the implications for inquiry which these pose. There are as yet no completely acceptable ways of classifying these grounds and drawing out their implications but one useful scheme has been suggested by Habermas [23]. He suggests that the grounds for justifying knowledge should be sought in the specification of the basic needs of communal existence. From the activities which are necessary for pursuing these needs flow the significant problems of inquiry. While one can and should refine and extend them, he proposes three major 'knowledge-interests' and three corresponding sciences.

Firstly, there is the knowledge-interest in 'technical control'. The knowledge appropriate for meeting this interest is that produced by the experimental (predominantly physical) sciences who attempt to provide through rigorously controlled procedures replicable knowledge of objectified processes. This type of knowledge currently dominates OR theory and practice. Secondly, there is knowledge-interest in 'interaction and communication'—the necessity for knowledge here is to sustain existence *as communal*. The knowledge appropriate for this interest is that produced by the 'hermeneutic' (predominantly social) sciences, who provide knowledge of social practices which can be consensually validated as true by social actors. The objectives of these sciences is to promote widespread understanding of these practices by the participants and not to provide a basis for their technical control. Hence, a dominant role for the methods of the physical sciences would be out of place here.

Whereas the second interest is in self-understanding, the third knowledge-interest is in 'self-control'—the necessity for knowledge here

is to understand the *origins* of social practices. Sustained communication and interaction is a necessary backdrop against which the pursuit of knowledge for technical control should proceed. Similarly, the sustained awareness of the origins of social practices is necessary as a backdrop against which they should be understood. The knowledge appropriate for this interest is produced by what Habermas terms the 'critical sciences' whose focus is on the way in which social and other structures (e.g. economic) subvert real needs.

At various points in their works writers on the nature of OR do in fact advocate the pursuit of all of these knowledge interests. However, confusion results when the interest of communication, interaction and self-control are pursued by the methods appropriate only to the interest of technical control. Thus, the investigation and resolution of conflict, which, if not predominantly, is, in its important aspects an activity of the hermeneutic and critical sciences, requires the application and development of *social-scientific* knowledge. Seen in this light Ackoff and Emery's attempt to develop knowledge of this social practice adequate for the interest of technical control was doomed to failure from its inception.

The general lesson to be learned from the failure of the philosophies of science to adequately define the nature of OR as a science—adequate to its whole system objective—is that philosophical analyses of ideal methods should be abandoned in favour of attempts to define and measure real knowledge interests and the knowledge which is necessary for their pursuit. If we follow the classification provided by Habermas there is clearly a much greater role for the social sciences in this model of the practice of OR than would be apparent from the works of the prominent theoreticians of OR.

SUMMARY AND CONCLUSION

None of the philosophies used by Ackoff, Beer and Churchman provide good grounds for compelling a client to accept the results of any investigation.

Positivism cannot interpret the phenomena to which it is restricted and pragmatism cannot justify its logic of inquiry without referring to a view of the whole system which exists before its inquiries are complete. Conventionalism/idealism cannot justify its ideas in their own terms and insofar as it attempts to make them objective in terms of a process of inquiry it has to refer implicitly to the real world. These processes therefore are merely a variant of the pragmatist's logic of inquiry which itself requires some external justification.

The failure of these philosophies highlights the basic weakness of the systems approach. This is the belief that true knowledge is transcendental in the sense of having some undeniable, fixed foundation which is independent of human purposes and social practices. Chasing this ideal has shown it to be a will-o'-the-wisp, and the game should now be abandoned to philosophers.

REFERENCES

1. ACKOFF RL (1960) Meaning, scope and methods of operational research. From *Progress in Operations Research, ORSA*, (ed. RL Ackoff), John Wiley, New York.
2. ACKOFF RL (1967) *Scientific Method, Optimizing Applied Research Decisions*, John Wiley, New York.
3. ACKOFF RL (1969) On the ambiguity of the researcher and the researched. From *The Place of Research in Social Choice*, Tavistock, London.
4. ACKOFF RL (1970) *A Concept of Corporate Planning*. John Wiley, New York.
5. ACKOFF RL (1972) Towards a system of systems concepts. From *Systems Behaviour* (eds J. Beishon and A Peters) Harper Row, London.
6. ACKOFF RL (1975) *Redesigning the Future*, John Wiley, New York.
7. ACKOFF RL (1975) A reply to the comments of Keith Chesterton, Robert Goodsman, Jonathan Rosenhead and Colin Thunhurst. *Ops Res. Q.* **26**(1), 96–99.
8. ACKOFF RL (1977) Optimisation + objectivity = opt out. *Eur. Jl Opl Res.* **1**, 1–7.
9. ACKOFF RL & RIVETT P (1963) *A Manager's Guide to Operational Research*, John Wiley, London.
10. ACKOFF RL & EMERY FE (1972) *On Purposeful Systems*, Tavistock, London.
11. BEER S (1966) *Decision and Control*. John Wiley, London.
12. BEVAN RG (1976) The language of operational research. *Opl Res. Q.* **27**, 1–8.
13. BEVAN RG & BRYER RA (1978) On measuring the contribution of OR. *Jl Ops Res. Soc.* **29**(5), 409–418.
14. CHECKLAND PB (1975) *Science and the systems paradigm*. Paper presented at 4th ICUS, New York.
15. CHURCHMAN CW (1961) *Prediction and Optimal Decision*, Prentice-Hall, Englewood Cliffs, New Jersey.
16. CHURCHMAN CW (1967) Wicked problems. *Mgmt Sci.* **14**(4), B–141–142.
17. CHURCHMAN CW (1968) *Challenge to Reason*, McGraw-Hill, New York.
18. CHURCHMAN CW (1970) Operations research as a profession. *Mgmt Sci.* **17**(2), B–37–53.
19. CHURCHMAN CW (1971) *The Design of Inquiring Systems*, Basic Books, New York.

20. CHURCHMAN CW & SCHAINBLATT AH (1965) The researcher and the manager: a dialectic of implementation. *Mgmt Sci.* **11**(4), B–69–87.
21. GIDDENS A (1970) "Power" in the recent writings of Talcott Parsons. From *Modern Sociology* (ed. P Worsley). Penguin, Harmondsworth.
22. GOULDNER AW (1959) Organizational analysis. From *Sociology Today* (ed. RK Merton). Harper Row, New York.
23. HABERMAS J (1972) *Knowledge and Human Interests*. Heinemann, London.
24. HOOS I (1972) *Systems Analysis and Public Policy: a Critique*, University of California Press, Berkeley and Los Angeles.
25. HUYSMANNS HBM (1970) *The Implementation of Operations Research*, John Wiley, New York.
26. KEAT R & URRY I (1975) *Social Theory as Science*. Routledge, London.
27. KEYNES JM (1936) *The General Theory of Employment, Interest and Money*, pp 383–4. Macmillan, London.
28. KOLAKOWSKI L (1972) *Positivist Philosophy: from Hume to the Vienna Circle*, Penguin, Harmondsworth.
29. KUHN T (1962) The structure of scientific revolutions. *Int. Encycl. Unified Sci.* Vol **II**, No 2, University of Chicago Press, London.
30. LOCKWOOD D (1970) Some remarks on 'the Social System'. From *Modern Sociology* (ed. P Worsley). Penguin, Harmondsworth.
31. LOUCH AR (1966) *Explanation and Human Action*, University of California Press, London.
32. LUKES S (1974) *Power*, Macmillan, London.
33. MILLS CW (1959) *The Sociological Imagination*. Penguin, Harmondsworth.
34. PIAGET J (1971) *Structuralism*, Routledge, London.
35. REX J (1963) *Key Problems in Sociological Theory*. Routledge, London.
36. RIVETT P (1974) Perspective for operational research. *Omega*, **2**(2), 225–233.
37. SMITH J (forthcoming) Why measure the contribution of OR? *Jl Opl Res. Soc.*
38. WAGNER HM (1971) The ABC's of OR. *Ops. Res.* **19**, 1259–1281.

ADDRESS FOR CORRESPONDENCE: *Dr RA Bryer, School of Industrial and Business Studies, University of Warwick, Coventry CV4 7AL, UK.*

ol. 7
979

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 3, pp. 233 to 240
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0601-0233$02.00/0

Scheduling and the Problem of Computational Complexity

JR KING

Imperial College, London

(*Received December* 1978)

Why is it that the problem of scheduling is so computationally difficult to solve? At last recent developments in modern mathematical complexity theory are providing some insights. The paper describes in essentially non-mathematical terms the computational technique known as the 'Branch and Bound Method'. This, the best general optimising technique available for scheduling, is also shown to have its limitations. It now appears that *efficient* computational and optimising algorithms are unlikely ever to be found for all except special cases of the general industrial scheduling problem. It seems that heuristic (rule-of-thumb) methods leading to approximate solutions are likely to offer the only real promise for the future.

INTRODUCTION

THE SEEMINGLY paradoxical situation whereby the industrial scheduling problem, even in its most elementary form, appears to be so difficult to solve and yet so easy to describe, has puzzled research workers in this area for many years. It has not, however, dampened enthusiasm or optimism that one day the problem will be cracked. Support for this view has derived from the fact that analytical optimising methods have been developed for a number of specific scheduling problems, involving in particular one or two machines.

Perhaps the most famous of these methods is that due to Johnson [6] who examined the flow scheduling problem where the number of jobs was unlimited but the number of machines was restricted to two. The term job is, in this context, understood to represent either a single item of product or a batch of products, and flow scheduling indicates that the order of processing of the jobs is the same on each machine. It is further assumed that the jobs are indivisible. Johnson's method provides an optimal solution by generating a schedule sequence that minimises the overall time to complete the schedule or as it is often called in the literature on this subject, the *makespan time*.

Johnson's method is stated below without proof and is applied to a simple illustrative example.

Johnson's method

(1) List $t(j, 1)$ and $t(j, 2)$ the processing times for n jobs labelled j, (where $j = A, B, C, \ldots$) on the two machines 1 and 2.

(2) Determine the minimum processing time amongst the $t(j, 1)$ and $t(j, 2)$.

(3) If it is a machine 1 time, put the corresponding job to the front of the sequence.

(4) If it is a machine 2 time, put the corresponding job to the rear of the sequence.

(5) Cross off both times for the job just allocated.

(6) Repeat from step 2 onwards with the reduced set of job times. The allocation of jobs at each stage by steps 3 and 4 being made working inwards (see example illustration problem) from the front or the rear of the job sequence so far determined.

(7) In the case of ties between the times for different jobs on the same machine, any order of processing of these jobs may be chosen. To provide an unambiguous tie breaking rule, we will always allocate such jobs in the same order that they were actually considered and if there is a choice between allocating to the front or back of the sequence, we will arbitrarily choose to always allocate to the front.

EXAMPLE PROBLEM

The processing times for a two machines and four jobs example problem are given in Table 1.

If Johnson's method is now applied to this problem, the stages that will be gone through are as illustrated in Table 2, from which it will be seen that the optimal (minimum makespan) job sequence is *C*, *B*, *D*, *A*. The corresponding Gantt chart representation of the complete schedule is shown in Fig. 1 and it will be seen that the minimum makespan time in this case is 39.

Johnson's method is an example of an *algorithm*. The term algorithm is defined by Chambers Twentieth Century Dictionary as 'a rule for solving a mathematical problem in a finite number of steps'. The importance of algorithms derives from the fact that they are necessary for the creation of programs for the solution of problems by computer.

A procedure can always be formulated which will consider each job in turn in each possible position in the sequence, so that a complete enumeration of alternatives is carried out. Such

TABLE 1.

Job j	Job processing times M/c 1 $t(j, 1)$	M/c 2 $t(j, 2)$
A	10	5
B	7	8
C	4	9
D	12	11

TABLE 2.

Stage	Sequence
i	C,....
ii	C,....A
iii	C, B...A
iv	C, B, D, A

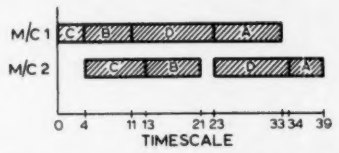


FIG. 1. *Gantt chart representation of the sequence C, B, D, A for the example problem.*

a procedure cannot be said to be mathematically elegant but nevertheless this too may be described as an algorithm. A total enumeration algorithm of this kind designed to generate sequences in, say, lexicographical (i.e. dictionary) order would, when applied to the example problem, produce the entire set of alternative sequences and in the order shown in the *tree* diagram of Fig. 2. The sequence in each case is read as the successive nodes linked by the branches running from the root (i.e. node all, indicating that all sequences are linked to this the root of the tree) to the highest level in the tree. Thus, for example, the first four sequences generated are read from left to right in the tree as *A*, *B*, *C*, *D*; *A*, *B*, *D*, *C*; *A*, *C*, *B*, *D* and *A*, *C*, *D*, *B*. The computed makespan times are shown alongside the last node in each sequence branch in Fig. 2. It is clear therefore that the optimal solution is *C*, *B*, *D*, *A* with makespan time of 39, the same as obtained by Johnson's algorithm.

What is also clear from a comparison of the two approaches is that Johnson's algorithm required 4 calculations whereas complete enumeration involved 24 (corresponding to the 24 separate sequences generated). For the general case of n jobs, Johnson's algorithm involves n calculations and complete enumeration requires $n! = n(n-1)(n-2)(n-3)\ldots 1$. The amount of computation is a function therefore, of what may be described as the 'size of the problem input' as represented by n and $n!$ respectively for the two approaches. If the computations are, as we assume, to be performed by computer, then the execution time of an algorithm which may be viewed as an effective measure of its efficiency, will, roughly speaking, be proportional to the number of calculations performed: which is in turn, as we have seen, a function of the size of the problem. It is useful therefore to differentiate between:

(1) Inefficient *exponential time* algorithms that give rise to execution times that increase rapidly with the size of the problem input,

Vol. 7
1979

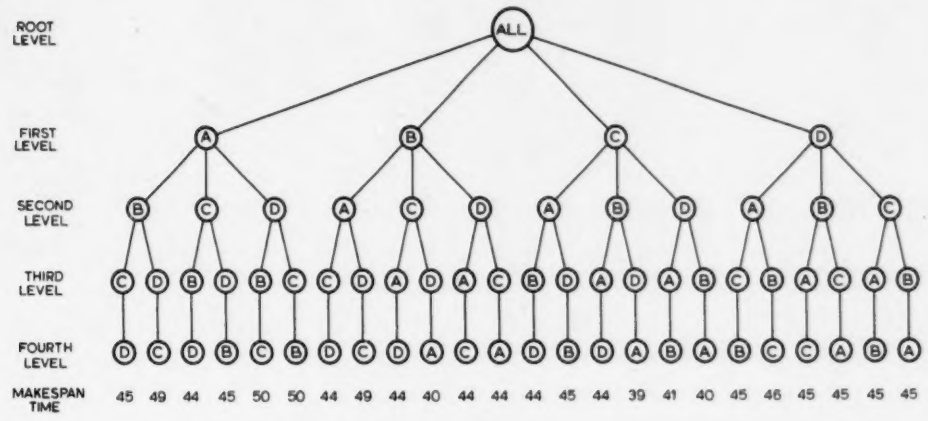

ROOT LEVEL
FIRST LEVEL
SECOND LEVEL
THIRD LEVEL
FOURTH LEVEL
MAKESPAN TIME
ALL
45 49 44 45 50 50 44 49 44 40 44 44 44 45 44 39 41 40 45 46 45 45 45 45

FIG. 2. *Complete tree of alternative sequences arranged in lexicographical order.*

or more specifically at what might loosely be termed 'exponential' rates of growth. Included in this exponential time group is the $n!$ case for the total enumeration algorithm just described and also cases of even faster growing execution times proportional to some number to the power of n, for some problems.

(2) Efficient *polynomial time* algorithms that result in execution times that increase relatively slowly with the size of the problem input, at a rate which in general can be represented by a polynomial function of n. That is a function involving one or more of the following terms n, n^2, n^3 *etc.* Usually it is sufficient in this context to define efficient algorithms as those which are bounded by a second degree polynomial time function. It is apparent therefore, that Johnson's algorithm is efficient.

Ever since Johnson published his algorithm, many attempts have been made and using a variety of optimising criteria, to discover efficient polynomial time bounded algorithms for the flow shop scheduling problem involving more than two machines and for the more general job shop scheduling problem where the various jobs may take different routes through the machines. With the exception of a few algorithms such as those developed for the single machine case by Smith [14] and Moore [13], a variation of the two machine problem by Jackson [5] and an extension by Johnson [6] of his two machine algorithm to a very restricted form of the three machines problem, relatively little progress has been made in discovering efficient algorthms for the scheduling problem.

It is clear that total enumeration algorithms are restricted as practical methods to problems of very small size. For problems of even modest size, such is the power of the exponential execution time requirement that although finite it would, if a computer were assigned to the task, take years of continuous computation before a solution would be generated.

THE BRANCH AND BOUND METHOD

If the total enumeration of alternatives is not possible, then perhaps a partial enumerative approach is? Such a technique is available and is usually referred to as the *branch and bound* method. The name derives from the fact that schedule sequences are built up by the generation of selected branches of the solution tree. The decision as to which branches should be generated and to what level, is determined by the calculation, at the appropriate nodes, of a lower bound on the partial sequence represented at that stage. When a sequence is generated which has a value less than or equal to any of the generated node lower bounds, then that schedule is the optimal solution.

To illustrate this important technique, a branch and bound algorithm is developed for application to the two machine flow shop and will be discussed in relation to the simple example problem described earlier. To avoid the usual mathematical complexities, the

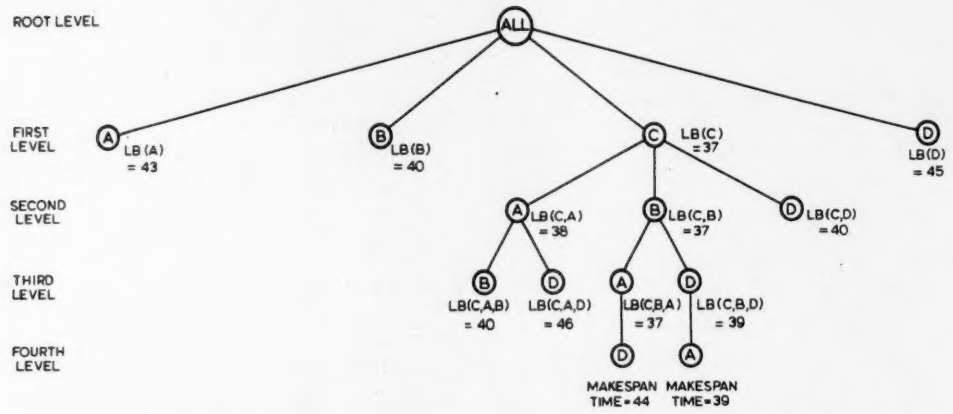


FIG. 3. *Tree generated by branch and sound algorithm in determining the optimal solution (sequence C, B, D, with makespan time = 39) to the example problem.*

bounding process will be described in terms of Gantt charts. The solution tree generated by the branch and bound method is shown in Fig. 3, this is built up progressively in a number of distinct stages in a manner about to be described.

Application to the example problem

Because the problem is a flow shop the order of processing for all jobs is the same, namely machine 1 first and machine 2 second, there will never be a delay between jobs being processed on the first machine. Any delays that may occur will arise on the second machine only and as a direct consequence of the particular job sequence selected.

Branchings from root node all to first level nodes A, B, C and D. The earliest time that the first job on machine 2 can be started is immediately it has been completed on machine 1. The shortest possible time necessary for processing all the jobs on machine 2 will arise if there are no delays, and will, therefore, be simply the sum of the processing times (note that this requires no specification of the processing sequence for the jobs).

A lower bound can therefore be established by adding the processing time of the first job on the first machine to the sum of the processing times of all the jobs on the second machine. Figure 4 illustrates the four lower bound values that result for the alternative choices of the first job to start the sequence. The results are transferred to Fig. 3 where the first level nodes *A, B, C* and *D* and the associated lower bound values 43, 40, 37 and 45 are entered.

Branchings from first level node C to second level nodes A, B and D. The next branching of the tree is chosen to take place from node *C* because it has the least lower bound value of any of the nodes so far generated. This may

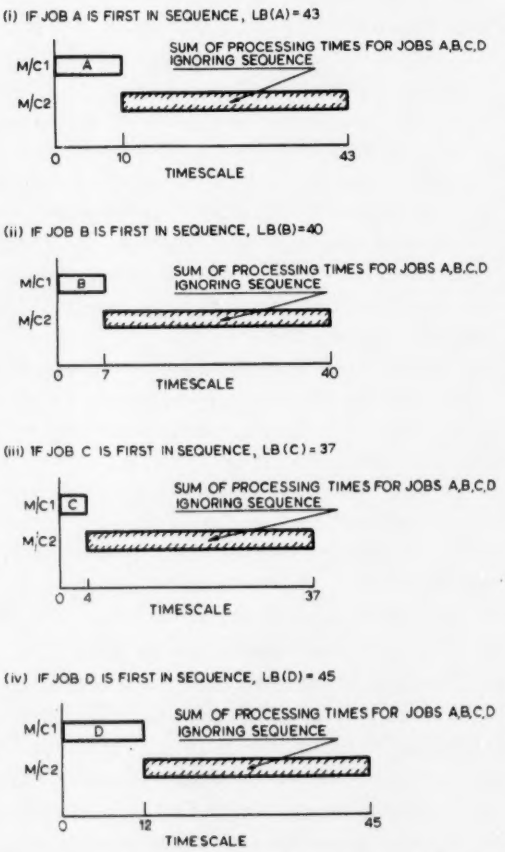


FIG. 4. *Lower Bound LB (?) when first job in sequence is to be assigned.*

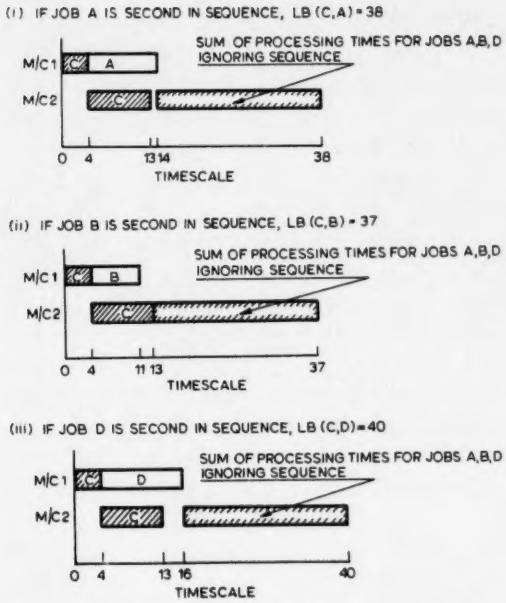

(i) IF JOB A IS SECOND IN SEQUENCE, LB (C,A) = 38
SUM OF PROCESSING TIMES FOR JOBS A,B,D IGNORING SEQUENCE
M/C1
M/C2
A
0 4 13 14 38
TIMESCALE
(ii) IF JOB B IS SECOND IN SEQUENCE, LB (C,B) = 37
SUM OF PROCESSING TIMES FOR JOBS A,B,D IGNORING SEQUENCE
M/C1
M/C2
B
0 4 11 13 37
TIMESCALE
(iii) IF JOB D IS SECOND IN SEQUENCE, LB (C,D) = 40
SUM OF PROCESSING TIMES FOR JOBS A,B,D IGNORING SEQUENCE
M/C1
M/C2
D
0 4 13 16 40
TIMESCALE

FIG. 5. *Lower Bound LB (C, ?) when job C is first in sequence and second job is to be assigned.*

not always necessarily be the best tree search strategy but will be the one adopted for illustrative purposes in this example. At this stage, therefore, the possibility of a complete schedule with makespan time of 37 (the lower bound value at node *C*) cannot be ruled out.

If job *C* is the first in the sequence, then the earliest start time for the remaining jobs on the second machine will be determined by the maximum of the earliest completion time of job *C* on machine 2 (i.e. the sum of the job *C* processing times on the two machines) and the earliest completion times for the second job in sequence on machine 1 (which is the sum of the processing times for this job and job *C* on machine 1).

A lower bound on the makespan time can be established by adding to the just computed earliest start time for the remaining jobs on machine 2, i.e. all jobs except job *C* already allocated), the sum of the processing times of these jobs on machine 2.

The lower bound values that result from the alternative partial sequences which all begin with job *C* but have respectively *A*, *B* or *D* as the second job in sequence, are illustrated in Fig. 5. These lower bound values are entered in Fig. 3 along with the associated second level nodes *A*, *B* and *D* which are reached by branches from node *C*.

Branchings from second level node B to third level nodes A and D. Level 2 node *B* is chosen for the next branching of the tree because it has the lowest lower bound value of all the pendant nodes (these are defined as the nodes at the ends of tree branches). The situation represented by the second level node *B* is a specified partial sequence designated by the branch all, *C*, *B* whereby jobs *C* and *B* are respectively processed first and second.

The earliest start time for the remaining jobs *A*, *D* on the second machine will be determined by the maximum of the earliest completion time of job *B* on machine 2 and the earliest completion time of the third job in sequence on machine 1 (i.e. the sum of the processing times of the first three jobs in sequence on machine 1).

A lower bound on the makespan time is determined by adding the earliest start time on machine 2 for the remaining jobs *A*, *D* and the sum of the processing times of these jobs on machine 2.

The resulting lower bounds on the makespan time for the two cases of *A* or *D* as the third job in sequence are illustrated in Fig. 6. The third level nodes *A* and *D* are now entered together with their associated lower bounds 37 and 39 in the tree of Fig. 3.

Branching from third level node A to fourth level node D. Branching from the lowest lower bound node means a single branch from node *A* to node *D*, since only job *D* is now left for assigning to the previous partial sequence. Thus, assigning job *D* to the last position in the sequence produces a complete schedule as illustrated in Fig. 7, with a makespan time of

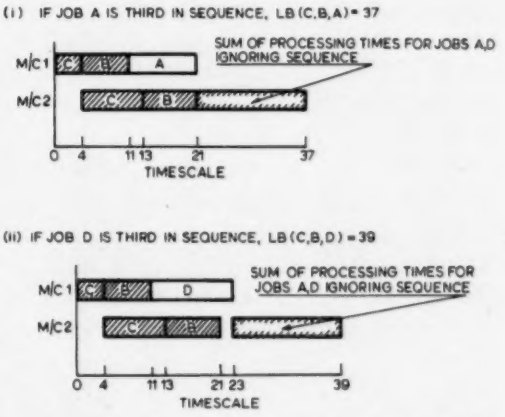

(i) IF JOB A IS THIRD IN SEQUENCE, LB (C,B,A) = 37
SUM OF PROCESSING TIMES FOR JOBS A,D IGNORING SEQUENCE
M/C1
M/C2
A
B
0 4 11 13 21 37
TIMESCALE
(ii) IF JOB D IS THIRD IN SEQUENCE, LB (C,B,D) = 39
SUM OF PROCESSING TIMES FOR JOBS A,D IGNORING SEQUENCE
M/C1
M/C2
D
0 4 11 13 21 23 39
TIMESCALE

FIG. 6. *Lower Bound LB (C, B, ?) when job C is first and B is second in sequence and third job is to be assigned.*

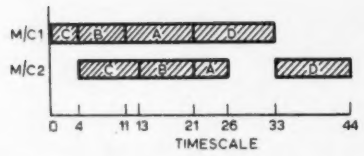


FIG. 7. *Complete schedule for sequence C, B, A, D with makespan time* 44.

44. This is larger than all bar one of the pendant node lower bound values, so that it is possible that another sequence with a lower makespan time exists.

Branching from second level node A to third level nodes B and D. Branching from the node with the lowest lower bound means branching from second level node *A* with lower bound value 38. At this stage the tree indicates a partial sequence of job *C* first and job *A* second. The earliest start time for the remaining jobs *B*, *D* on the second machine is specified by the maximum of the earliest completion time of job *A* on machine 2 and the earliest completion time of the third job on machine 1.

This is illustrated in Fig. 8, where it will be seen that adding the sum of the machine 2 processing times of the remaining jobs *B*, *D* to the earliest start time on the same machine, provides lower bounds on the makespan time for the two cases where the third job in sequence is *B* or *D*. These lower bounds are respectively 40 and 46, and are entered in the tree of Fig. 3 alongside the third level nodes *B* and *D*, the ends of the branchings from second level node *A*.

Branching from third level nodel D to fourth level node A. The lowest lower bound value is now 39 and is associated with the pendant

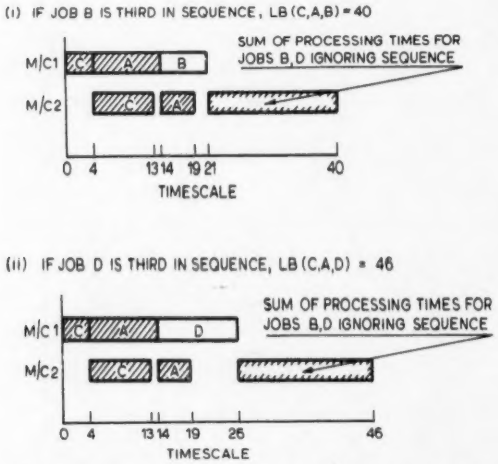

(i) IF JOB B IS THIRD IN SEQUENCE, LB (C,A,B) = 40
SUM OF PROCESSING TIMES FOR JOBS B,D IGNORING SEQUENCE
M/C1
M/C2
B
0 4 13 14 19 21 40
TIMESCALE
(ii) IF JOB D IS THIRD IN SEQUENCE, LB (C,A,D) = 46
SUM OF PROCESSING TIMES FOR JOBS B,D IGNORING SEQUENCE
M/C1
M/C2
D
0 4 13 14 19 26 46
TIMESCALE

FIG. 8. *Lower bound LB (C, A, ?) when job C is first and A is second in sequence and third job is to be assigned.*

node *D* at level 3 in the tree shown in Fig. 3. There is only one branch possible and that is from *D* to the level 4 node *A*, which corresponds to the adding of the final job *A* to complete the sequence *C, B, D, A*. This schedule has already been illustrated in Fig. 1 and has a makespan time of 39.

Testing for optimality. Reference to Fig. 3 shows that sequence *C, B, D, A* has a lower makespan time than the other complete sequence *C, B, A, D* which has a value 44. The makespan time of 39 for *C, B, D, A* is also lower than any of the lower bound values for the other pendant nodes developed in Fig. 3. It is, therefore, impossible for any other complete schedule to be generated by branchings from these pendant nodes that would have a lower makespan time value. This is clearly so, since the lower bound value may increase but never decrease in a branching from one node level to the next lower level. Thus, the sequence *C, B, D, A* must be the optimal solution.

This simple illustration shows how a branch and bound algorithm may be devised which enables an optimal solution to be determined by a process of selective partial enumeration of the tree of alternative solutions. The algorithm used is not intended as a rival to Johnson's algorithm but has been developed here as a vehicle for explaining the basic ideas behind the branch and bound method.

COMPUTATIONAL COMPLEXITY

The nature of problem complexity

The early 1930's theoretical work by the British mathematician AM Turing [15] on the properties of algorithms has recently been used and extended, in an attempt to obtain a fundamental understanding of the complexity of mathematical problems. An important discovery by Turing was that all imaginable problems of mathematics could be divided into two mutually exclusive categories: (i) those for which it can be shown that no algorithm can be written, which are therefore in a formal sense provably unsolvable, and (ii) those which can be solved by algorithms.

We are concerned here with problems in the second category and the following discussion is applicable only to this group.

Modern computational complexity theory as developed by Cook [4] and Karp [7], desig-

Vol. 7
1979

nates that all problems[1] that can be solved by enumerative exponential time algorithms, are members of a class *NP*. Problems with known polynomial time algorithms are said to be members of a class *P*. It is clear that *P* is a subset of *NP* since some problems, such as for example the two machine flowshop that was considered earlier, have both polynomial and exponential time algorithms. Where both types of algorithm exist, obviously the efficient polynomial time one would be selected.

It cannot be ruled out that polynomial time algorithms may exist for problems in *NP* (and not already included in *P*). All that can be said is that in spite of much searching, if they exist, they have not as yet been found. Indeed, it is possible that polynomial time algorithms may exist for all problems in *NP*, which would mean that *NP* and *P* are identical; mathematicians are now generally agreed however, that this is extremely unlikely as all the indications are against it.

Cook [4] has shown that it is possible to formally identify certain combinatorial problems within the class *NP* which he terms *NP*–complete. These problems have been shown to have the remarkable property of *reducibility*, which essentially means that if a polynomial time algorithm could be found for any one of them, then polynomial time algorithms could be constructed from it for solving all the other problems in the class *NP*–complete.

NP–complete problems constitute a large class of notoriously difficult combinatorial problems for which no efficient polynomial time algorithms are available. The significance of formal identification of a problem as being *NP*–complete is that it represents an acceptance that enumerative methods such as branch and bound should be used since it is extremely unlikely that any better optimisation method will ever be found to exist.

Excluding certain specific situations involving one and two machines, some of which were mentioned earlier, both the flow scheduling and the more general job-shop scheduling problems have been shown to be *NP*–complete. Indeed, Lawler [8] has shown that even the one machine problem becomes *NP*–complete, if the problem is specified as that of determining the minimum mean flow time schedule, when subject to arbitrary precedence constraints. The scheduling problem in which the job set up times depend upon the job sequence used, is also *NP*–complete. This is in fact the production equivalent of the famous 'travelling salesman' problem, which is to determine the minimum distance tour of a number of cities for a salesman who is required to visit each city once only.

Branch and bound algorithms for the three or more machines flow shop scheduling problem and the job shop scheduling problem, have been developed by Lomnicki [11], Brown and Lomnicki [3], McMahon and Burton [12], Bestwick [1], Brooks and White [2] and many other authors. A branch and bound algorithm has also been developed for the travelling salesman (sequence dependent set up time) problem by Little *et al.* [10].

The efficiency of a branch and bound algorithm depends on how restricting are the bounds that are used. Generally speaking, the tighter the bounds, the more efficient the algorithm because less of the tree needs to be searched. Offset against this may be a loss of efficiency because the tighter bounds are more complicated and require more time to compute.

The limitation of branch and bound methods

Branch and bound methods are, as has been demonstrated, very much more efficient than complete enumeration. Nevertheless, no matter how well they are constructed, they also are exponential time algorithms and although they allow us to obtain optimal solutions, there is necessarily a finite limit to the size of problem that can be solved. This is not a rigid constraint since it will depend upon the amount of computer time that the user is willing to devote to the problem.

It is clear that sooner or later a problem size is encountered, for which it must be concluded that the computer time required is excessive. Such is the power of the exponential time requirement, that point is reached sooner

[1] The classes *P* (polynomial) and *NP* (non-deterministic polynomial) are drawn from the field of language recognition and constitute problems that are respectively solvable by deterministic and non-deterministic Turing machines. A Turing machine is simply a form of hypothetical computer which Turing used as a mechanism for studying various forms of mathematical problems. Turing never actually constructed such machines, they were and still are mathematical devices useful to the study of mathematical complexity.

rather than later. Branch and bound methods extend the size of optimising problems that can be solved, they are the best methods that we have for *NP*–complete problems but as has been demonstrated, their usefulness is also limited. Lewis and Papadimitriou [9], in an interesting review of the efficiency of algorithms, make the point that although for problems of small size the polynomial time function may be greater than the exponential time function, there will always be a size of problem at which the exponential function will overtake the polynomial, whatever its form. What is apparent is that the only type of algorithm that is really useful for all problems, including those of large size, is that which gives rise to polynomial execution time, in other words an efficient algorithm.

The conclusion must be drawn that there seems little point in expending further effort in seeking optimal solutions to the very large scheduling problems of the kind frequently encountered in many practical situations, they are certain, for at least the forseeable future, to be beyond the available computational capabilities. For such problems it would appear that we must concentrate on constructing heuristic polynomial time algorithms that seek approximate solutions that are satisfactory, even though they may not be optimal.

REFERENCES

1. Bestwick PF & Hastings NAJ (1976) A new bound for machine scheduling. *Opl Res. Q.* **27**, 479–487.
2. Brooks GH & White CR (1965) An algorithm for finding optimal or near optimal solutions to the production scheduling problem. *Jl Ind. Engng*, **16**, 34–40.
3. Brown APG & Lomnicki ZA (1966) Some applications of the branch and bound algorithm to the machine scheduling problem. *Opl Res. Q.*, **17**, 173–186.
4. Cook SA (1971) *The complexity of theorem-proving procedures.* Proc. 3rd Annual ACM Symp. Theory Comput. 151–158.
5. Jackson JR (1956) An extension of Johnson's results on Job Lot scheduling. *Nav. Res. Logist. Q.* **3**, 201–203.
6. Johnson SM (1954) Optimal two- and three-stage production schedules with set up times included. *Nav. Res. Logist. Q.* **(1)**, 61–68.
7. Karp RM (1972) Reducibility among combinatorial problems. From *Complexity of Computer Computations* (Eds Miller RE & Thatcher JW), Plenum Press, New York, 85–103.
8. Lawler EL (1976) Sequencing jobs to minimize total weighted completion time subject to precedence constraints. *Annals of Discrete Mathematics.* **2**, 75–90.
9. Lewis HR & Papadimitriou (1978) The efficiency of algorithms. *Scient. Am.* **238**(1), 96–109.
10. Little JDC, Murty KG, Sweeney DW & Karel C (1963) An algorithm for the travelling salesman problem. *Ops Res.* **11**(6), 972–989.
11. Lomnicki ZA (1965) A branch and bound algorithm for the exact solution of the three-machine scheduling problem. *Ops Res. Q.* **16**, 89–100.
12. McMahon GB & Burton PG (1967) Flow-shop scheduling with the branch and bound method. *Ops Res.* **15**, 473–481.
13. Moore JM (1968) An *n* job, one machine sequencing algorithm for minimising the number of late jobs. *Mgmt Sci.* **15**, 102–109.
14. Smith WE (1956) Various optimisers for single stage production. *Nav. Res. Logist. Q.* **3**(1) and (2), 59–66.
15. Turing AM (1937) On computable numbers, with an application to the Entscheidungs problem. *Proc. London Mathematical Soc.* (2), **42**, 230.

ADDRESS FOR CORRESPONDENCE: *JR King, Esq, Department of Management Science, Imperial College, Exhibition Road, London, SW7 2BX.*

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 3. pp. 241 to 247
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0601-0241$02.00/0

Revision of Consumer Intention: Two Models

JAMES W GENTRY

Oklahoma State University, Stillwater

JAMES SHANTEAU

Kansas State University, Manhattan

(*Received July* 1978; *in revised form November* 1978)

Two models of opinion revision, a Bayesian model and a proportional change model, are compared on their abilities to describe subjects' revisions of purchase intentions upon the receipt of consumer information. Both models were found to predict the revisions reasonably well, but the proportional change model performed significantly better.

ol. 7
979

INTRODUCTION

VARIOUS models of probability revision for inference judgments have been considered in recent literature (for a comprehensive review, see [20]). These models tend to originate either from Bayesian or information-processing considerations. The Bayesian approach uses Bayes' theorem to describe the normative or optimal combination of sample information with prior opinions to form a revised opinion. In comparison, information processing models are purely descriptive and range from simple additive models, including regression models, to more complex weighted averaging models.

Generally, the two types of models have been used with different purposes in mind. The usual purpose in Bayesian studies has been to compare subject's probability revisions with optimal standards derived from Bayes' theorem. Much of the Bayesian literature has been concerned with the causes of 'conservatism', i.e. the failure of subjects to extract the full diagnosticity from sample information. In contrast, the usual purpose behind information processing studies has been to describe the psychological processes involved in inference judgment. This has led to analyses of the combination rules used by subjects to arrive at a judgment.

It has frequently been suggested that probability revision models might be useful in studying consumer information acquisition and information processing [10, 13]. Almost without exception, however, marketing studies discussing the revision of inferences have been based on a normative approach [4, 8, 9, 12], with most of these studies making use of abstract tasks. For instance, Green [9] used three urns of pokerchips to investigate 'brand switching'. Generalizing from such contrived situations to consumer information processing seems questionable at best.

Similarly, most information processing research dealing with probability revision has also been based on the 'bookbags-and-pokerchips' task. This task, popularized in early Bayesian research, calls for an inference about which of two bags of pokerchips produced a particular sample of chips. Unfortunately there has been little effort to generalize any of the previous information-processing results to a consumer setting.

The purpose of this study was to compare Bayesian and information processing models of consumer decision making. Most previous

research has used scenarios which allow comparison of the subject's judgments with an optimal standard. In most real-world tasks, however, theoretical standards are not available. In this study, the Bayesian and information-processing models will be compared on their ability to describe consumers' revisions of purchase intentions.

MODIFIED BAYESIAN MODEL

Bayes' theorem dictates that optimal manner in which prior information and sample information should be combined to provide the posterior estimates of the relative uncertainties. Let $P(H_i)$ represent the probability that Hypothesis i (for example, that store i has the best buy for the respondent) is true prior to the presentation of sample information. Let $P(H_i|D)$ represent the posterior probability following the information, and $P(D|H_i)$ represent the conditional probability of the sample information given that the hypothesis is true. Then the normative probabilities are specified by Bayes' theorem:

$$P(H_i|D) = \frac{P(H_i)P(D|H_i)}{\sum_{j=1}^{n} P(H_j)P(D|H_j)}. \quad (1)$$

While the normative properties of the Bayesian model are beyond dispute, the descriptive properties have been questioned [17–21]. Other writers, however, have suggested that the Bayesian model may still be of some usefulness as a descriptive model of human decision making [14, 16].

In the present study, the Bayesian model was modified so that surrogate measures were used instead of conditional probabilities. Two types of surrogate measures described below were used in equation 1 to try to predict the posterior probabilities.

Surrogate likelihood measures

In the abstract pokerchips paradigm, it is possible to ask subjects to give the conditional probabilities (or likelihoods) of observing a sample given the various bookbags. For instance, subjects can be asked "What is the probability that you would see a sample of four red and three white pokerchips if a bookbag with 70 red and 30 white chips is sampled?" In contrast, the comparable question in a consumer setting is more difficult, e.g. "What is the probability that you would see information such as this given that Store *A* has the best buy?" Questions similar to the latter were asked of pilot subjects. However, their responses and subsequent comments revealed that they were not able to understand the question satisfactorily. In fact, it was necessary in an earlier study [6] to make the consumer-oriented task almost as abstract as the bookbags and pokerchip paradigm in order to assess likelihoods in a consumer setting.

Instead of conditional probabilities, it was decided to ask subjects to supply 'surrogate' measures of probability. Two surrogate measures were used: an attitude surrogate and an agreement surrogate.

The *attitude* surrogate asked the respondent to rate the impression that he received of the alternatives under consideration based on the sample information. Assuming that two subjects had equal prior probabilities of purchasing at Store *A*, a subject who said that the sample information gave him a favorable impression of Store *A* was expected to assess a higher current (posterior) probability of purchasing there than would a subject who said that the sample information gave him an unfavorable impression of the store. This fits intuitively with what Bayes' Theorem would dictate since large likelihoods would presumably lead to more favorable impressions. Operationally, the subjects were asked to rate their attitude toward each store on the basis of the sample information alone. This rating was obtained on a nine-point scale ranging from extremely unfavorable to extremely favorable.

The *agreement* surrogate asked the respondent to assume that each alternative was true and then to rate the agreement between that assumption and their impression of the sample information. Intuitively, a large likelihood would mean that there is a strong agreement between what the information says and the truth of the hypothesis under consideration, and a small likelihood indicates weak agreement. Operationally, subjects were asked to assume that a particular store had the best buy for them, and then to assess the amount of agreement between this assumption and the sample information. These assessments were obtained on nine-point scales ranging from ex-

Vol. 7
1979

tremely strong disagreement to extremely strong agreement.

The validity of these surrogates was investigated in situations in which the assessment of the likelihoods themselves was practical [5]. Results showed a significant relationship between the surrogate likelihoods and the assessed likelihoods in a consumer-oriented task.

PROPORTIONAL CHANGE MODEL

The particular information processing model examined in this study is the proportional change model described by Anderson [1, 2]. According to this model, the current attitude or opinion is a weighted average of the prior opinion and the sample information. This model can be stated as

$$R_1 = (1 - w)R_0 + (w)s. \quad (2)$$

where R_0 is the prior opinion, R_1 is the posterior opinion, s is the evaluation of the stimulus from the sample information, and $(1 - w)$ and (w) represent the relative weightings of the two information sources.

While the proportional change model was originally applied by Anderson to attitude change, it was extended to sequential inference judgments by Shanteau [22–24, 26]. An important prediction of this model is that sample information and prior opinions should combine by averaging. This prediction is in contrast to the multiplying combination rule predicted by the Bayesian model. The averaging prediction has been supported in the bookbag paradigm [19] and in a numerical inference task [11]. Other predictions of this model, such as additivity and order effects, have also received some support [18, 22].

In the present study, the attitude surrogate measure was used to provide estimates of the sample information value. In addition, subjects were also asked to estimate the relative weighting of prior opinions and sample information. These values were used in equation 2 to try to predict the posterior opinions. This means that a direct comparison is possible between equations 1 and 2 on their abilities to predict the subjects' responses.

METHOD

Subjects and stimuli

Sixty-five paid female undergraduates at Indiana University participated in the study. The task selected involved shoe purchase in Bloomington, Indiana. Shoes represent a fairly high involvement product category for females, and all subjects indicated that they planned to purchase shoes in Bloomington in the next three months.

The subjects were first asked to consider the type of shoes that they would be purchasing next. Then they were handed a deck of cards containing the names, addresses, and pictures of nearly all the stores in Bloomington which sell shoes. From this list they were asked to choose the stores at which they would most likely consider shopping. If there was a store in which they were interested, but which was not listed, they could also include that store in their set.

Procedure and instructions

The subjects were asked to (1) make prior probability and attitude assessments, (2) receive information about the shoe stores, (3) assess posterior probabilities and attitudes, (4) assess the surrogate likelihoods, and (5) determine the weight of the sample information compared to their prior information. They were then shown a second source of information and steps (3), (4), and (5) were repeated. The specific steps mentioned above will now be discussed in greater detail.

The subjects were asked to assess the probability that each store in the set (the alternate 'all others' was added to insure that the alternatives were exhaustive) had the best buy for them $[P(H_i)]$. They responded by using a 'line segment' approach, which required them to circle line segments according to how likely it was that they would find their best buys at a given store. Pilot studies indicated this technique was much simpler for the subjects to use than direct probability assessments.

After circling the line segments, they handed the sheet to the experimenter, who quickly normalized and wrote down the corresponding probabilities and then handed the sheet back to the subjects. They were told that, if these probabilities did not accurately represent their

feelings, they were free to change the probabilities either by going through the line segment approach again or by changing the numerical probabilities directly (as long as they still summed to one). Next they were asked to give their prior attitudes (R_0) toward the stores that had been chosen, using nine-point scales ranging from "Extremely Unfavorable" to "Extremely Favorable".

They were then shown two sources of consumer information consisting of shoe advertisements from Bloomington newspapers and a purported publication of a local consumer action group on the Bloomington shoe market. The order in which the subjects saw the information was counterbalanced over sessions.[1] These sources contained accurate product line and price information on almost all retail outlets carrying women's shoes in Bloomington. The information sources were designed to be as diagnostic as possible so that the opinion change taking place in the laboratory would be meaningful. However, there was no inducement for the subjects to change their opinions and four out of the 65 subjects did not choose to revise their opinions on one or both of the iterations. Having seen the first source of consumer information, they were asked to assess the posterior probability that each alternative store had the best buy for them [$P(H_i|D)$], as well as their posterior attitudes (R_1) toward the stores.

Finally, they were asked to give measures of the impact of the information using the attitude and agreement likelihood surrogate measures. These values were used as estimates of the subjects' likelihoods in the Bayesian model [$P(D|H_i)$]; the attitude surrogate also served as information value in the proportional change model (s). The subjects were also asked to weigh the importance of the most recent information (w), as compared to their prior knowledge, in the determination of their final opinions of the stores. The entire procedure was repeated after a second source of information was shown to the subjects.

RESULTS

The mean number of stores listed by the subjects was 4.38 stores, with a median of 4 stores. Some of the subjects chose as few as 3 stores while one selected 10. In total, the 65 subjects made 350 revisions after being shown the first sample information and 350 more after seeing the second information. For each subject, the predictions of the two models were calculated using equations (1) and (2).

Correlational analyses

The correlations[2] between the assessed posterior measures and the models' predictions were as follows: first revision—0.74 for the modified Bayesian model using the attitude surrogate, 0.74 for the modified Bayesian model using the agreement surrogate, and 0.79 for the proportional change model; second revision—0.81 for the modified Bayesian model using the attitude surrogate, 0.82 for the modified Bayesian model using the agreement surrogate, and 0.83 for the proportional change model. Thus the results indicate that the proportional change model was somewhat better than the modified Bayesian models. Using the Fisher r to z transformation, however, no significant differences were found between the correlations for the models.

The correlations between the models' predictions and the posterior opinions were compared with the correlations between the prior and posterior measures (first revision, 0.57 for the attitudes and 0.60 for the probabilities; second revision, 0.65 for the attitudes and 0.74 for the probabilities; both revisions,[3] 0.39 for the attitudes and 0.51 for the probabilities). In each case, the correlations between the models' predictions and the assessed opinions were significantly higher than the correlations between

[1] Pretesting had indicated that both information sources would stimulate posterior opinions that were different from the prior ones. Analysis of the impact of the sources in the current study found that the first source shown the subjects (regardless of its content—advertisements or objective survey) resulted in more revision as measured by the mean absolute deviation from prior to posterior assessment. However, the amount of revision following the second source was not significantly smaller than the earlier revision.

[2] The posterior attitude measures were normalized for each subject so that differential response-scale tendencies could be removed before the measures were aggregated over subjects. Next the assessed posterior measures were correlated with the posterior measures derived from the models. It should be pointed out that the different sets of data are not completely independent since the impact of each subject on the overall correlation results is a function of the number of stores selected by her.

[3] "Both" means that the prior measurements of the first revision were correlated with the posterior measures of the second revision.

Vol. 7
1979

TABLE 1. MEAN RATIOS OF THE PREDICTED POSTERIORS TO THE ASSESSED POSTERIORS

Iteration	Modified Bayesian Model Attitude surrogate	Modified Bayesian Model Agreement surrogate	Proportional change model
First	1.05 (SD = 0.17)	1.07 (SD = 0.14)	0.98 (SD = 0.08)
Second	0.96 (SD = 0.05)	0.99 (SD = 0.08)	1.01 (SD = 0.09)

the prior and posterior measures. This finding indicates that marked revisions were taking place and that the models were able to account for the revisions quite well.

Analyses of model discrepancies

A shortcoming of comparing predicted and observed responses through correlations is that this analysis is not always sensitive to deviations from a model [3]. Further analyses of the data concentrated on the deviations of the predicted posteriors from the assessed posteriors, i.e. the discrepancies from the models. In order to look at the discrepancies on a comparable scale, the models' predictions were divided by the corresponding observed responses.[4] The mean predicted-assessed ratio was determined for each subject for each model, and then the mean ratios were aggregated over all subjects. If the models' predictions were consistent with the subjects' estimates, then we would expect a mean ratio of 1.0. The results are shown in Table 1.

When the mean ratios were compared to a value of 1.0, the predictions from the modified Bayesian model using the attitude surrogate were significantly different from the subjects' estimates ($p < 0.025$ in the first revision and $p < 0.01$ in the second). The predictions of the Bayesian model using the agreement surrogate were significantly different from the subjects' estimates for the first revision ($p < 0.001$) only; there were no significant differences between the model's predictions and the subjects' estimates in the second revision ($p > 0.25$). The differences for the proportioned change model were nonsignificant for both the first ($p > 0.05$) and second revision ($p > 0.25$).

Comparison between models

The final analysis involved a direct comparison between the models, again using the ratios of the predicted posterior measures to the subjects' assessed posterior measures. The absolute deviation of the ratios from 1.0 was computed for each estimation, aggregated over the stores for each subject, and then the mean deviations were averaged across subjects. The three models were compared in a pairwise fashion for both the first and second revisions with the results shown in Table 2.

The proportional change model was found to predict significantly better ($p < 0.001$) than either of the modified Bayesian models. The modified Bayesian model with the attitude surrogate was found to predict significantly better ($p < 0.001$) than the modified Bayesian model with the agreement surrogate. Thus, all three sets of analyses were consistent in finding a slight superiority of the proportional change model. Further, the modified Bayesian model with the attitude surrogate appeared slightly superior to the model with the agreement surrogate.

DISCUSSION

There are two major implications of this study that deserve elaboration. The first concerns the comparison between Bayesian and information-processing models. The second concerns the larger issue of probability revision in consumer and marketing research:

Comparison between models

As noted by Rappoport and Wallsten [15], the question in comparing decision-making models should not be: which model is right or wrong? Instead, the question should be: in what situation(s) is each model appropriate?

[4] The models' discrepancies were also investigated using such measures as the percentage error in the predictions and the ratio of the mean absolute deviation between the models' predictions and assessed posterior measures to the mean absolute deviation between the assessed prior and posterior measures. The relative performances of the different models were similar in all forms of analysis.

TABLE 2. COMPARISONS OF THE MODELS USING PAIRED OBSERVATIONS OF THE ABSOLUTE DEVIATIONS OF THE PREDICTED-ASSESSED RATIOS FROM 1.0

Revision	Models compared (superior model listed first in each case)	Difference in mean deviations	Standard deviation of difference
First	Proportional change Modified Bayesian (attitude surrogate)	0.21*	0.18
First	Proportional change Modified Bayesian (agreement surrogate)	0.29*	0.19
First	Modified Bayesian (attitude surrogate) Modified Bayesian (agreement surrogate)	0.08*	0.19
Second	Proportional change Modified Bayesian (attitude surrogate)	0.09*	0.09
Second	Proportional change Modified Bayesian (agreement surrogate)	0.18*	0.16
Second	Modified Bayesian (attitude surrogate) Modified Bayesian (agreement surrogate)	0.09*	0.18

*$p < 0.001$

Therefore, the properties of a situation should serve as a guide in selecting the most appropriate model.

The results from this study support the suggestion that both the Bayesian model and the proportional change model may be applicable in certain situations. The proportional change model appears to provide a generally useful approach to the study of consumer opinion change. The model consistently performed better than did either of the modified Bayesian models. Thus, we recommend that those interested in consumer opinion change might well explore the proportional change model further.

The modified Bayesian models did an adequate job of describing consumers' revision of probabilities, though inferior to the proportional change model. It is not surprising that a normative model would describe subjects' performances less well than a descriptive model. However, the Bayesian approach is sufficiently versatile to handle several different data generating functions and the influence revisions associated with them. While this study does not strongly support the use of the Bayesian model in a descriptive manner, the results certainly do not preclude this. Also, the Bayesian approach appears to be more flexible in regard to surrogate measures than has been previously considered.

If relevant consumer tasks can be constructed such that subjects can provide valid estimates of the necessary prior, posterior, and likelihood probabilities, then it may be possible to delineate 'good' consumer decisions from 'poor' ones. Consumer advocates frequently imply that 'good' decisions can be defined for consumers as a group. However, it seems more reasonable to talk about a 'good' decision for an individual consumer, given his/her utility function and situational constraints. Further work on the normative Bayesian approach may provide the flexibility needed to provide operational definitions of such 'good' consumer decisions.

Probability revision and consumer opinion change

The use of probabilistic purchase intentions has been hampered by the difficulty in obtaining valid probability assessments. The measurement problem increases when we try to investigate the impact of marketing stimuli on those purchase intentions. Research on probability revision should eventually result in an improved understanding of consumer opinion change. However, much of the probability revision research to date has dealt with abstract scenarios such as the bookbag and pokerchip paradigm. This research offers little to those interested in marketing applications.

The development of consumer relevant tasks, as pointed out here and elsewhere [6], is not without complications. The use of surrogate measures, for instance, needs to be examined in greater detail. In order to make the consumer task realistic in most purchase situations, the subject will have to be confronted with multiple alternatives instead of the two-alternative task common to much of the probability revision research. This introduces added

problems in probability measurement since the task becomes more difficult for the subject. Also problems in model testing may arise because the discrete probability distributions are linearly dependent. Fortunately, the present results revealed that various approaches to testing were generally consistent.

One further concern about the study of opinion change in the laboratory is that the change may be artificially induced and hence inherently unstable. However, a follow-up [7] to this study found that the posterior measures predicted actual purchase behavior better than did the prior measures.

REFERENCES

1. ANDERSON NH (1971) Integration theory and attitude change. *Psychol. Rev.* **78**, 171–206.
2. ANDERSON NH (1974) Information integration theory: a brief survey. From KRANTZ DH, ATKINSON RC, LUCE RD & SUPPES P. *Contemporary Developments in Mathematical Psychology*, 236–305. Freeman, San Francisco.
3. ANDERSON NH & SHANTEAU J (1977) Weak inference with linear models. *Psychol. Bull.* **84**, 1155–1170.
4. BASS FM (1963) Marketing research expenditures: a decision model. *Jl Bus.* **36**, 77–90.
5. GENTRY JW (1975) *The measurement of subjective likelihoods in relevant situations: the use of likelihood surrogates.* From HALL RW, Proc. Midwest AIDS Conf. Indianapolis, April 3–5, 277–280.
6. GENTRY JW & O'BRIEN TV (1978) Conservatism in Bayesian information processing. *Omega*, **6**(3), 280–281.
7. GENTRY JW & SUMMERS JO (1976) Are laboratory induced revisions in purchase intentions meaningful? *Jl Bus. Res.* **4**, 177–186.
8. GREEN PE (1962) Bayesian decision theory in advertising. *Jl Advertg Res.* **2**, 33–41.
9. GREEN PE (1966) Consumer use of information. From Newman J, *On Knowing the Consumer*, 67–80. John Wiley, New York.
10. GREEN PE, FITZROY PJ & ROBINSON PJ (1968) Experimental gaming in the economics of information. From Bass FM, King CW & Pessemier EA, *Applications of the Sciences in Marketing Management*, 333–356. John Wiley, New York.
11. LICHTENSTEIN S, EARLE TS & SLOVIC P (1975) Cue utilization in a numerical prediction task. *Jl Expl Psychol.: Hum. Percept. and Perf.* **104**, 77–85.
12. MOSKOWITZ H (1971) *Conservatism in group information processing behavior under varying management information systems.* Paper presented at Third Res. Conf. on Subjective Probability, Utility and Decision-Making, Brunel University, London, UK.
13. PETERS WS (1964) Utility, uncertainty and the consumer buyer. From Cox R, Alderson W & Shapiro SJ, *Theory in Marketing*, 254–269. Homewood, Illinois.
14. PETERSON CR & BEACH LR (1967) Man as an intuitive statistician. *Psychol. Bull.* **68**, 29–46.
15. RAPPAPORT A & WALLSTEN TS (1972) Individual decision behavior. *A. Rev. Psychol.* **23**, 131–176.
16. SAVAGE LJ (1964) The foundations of statistics reconsidered. From Kyburg HE & Smokler HE, *Studies in Subjective Probability*, 175–189. John Wiley, New York.
17. SHANTEAU J (1970) An additive model for sequential decision making. *Jl Expl Psychol.* **85**, 181–191.
18. SHANTEAU J (1972) Descriptive versus normative models of sequential inference judgment. *Jl Expl Psychol.* **93**, 63–68.
19. SHANTEAU J (1975) Averaging versus multiplying combination rules of inference judgment. *Acta Psychol.* **39**, 83–89.
20. SLOVIC P, FISCHHOFF B & LICHTENSTEIN S (1977) Behavioral decision theory. *A. Rev. Psychol.* **28**, 1–39.
21. TROUTMAN CM & SHANTEAU J (1977) Inference based on nondiagnostic information. *Orgnl Behavl and Hum. Perf.* **19**, 43–55.
22. WALLSTEN TS (1972) Conjoint-measurement framework for the study of probabilistic information processing. *Psychol. Rev.* **79**, 245–260.

ADDRESS FOR CORRESPONDENCE: *James W Gentry, Esq. College of Business, Oklahoma State University, Stillwater, OK 74074, USA.*

BLANK PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 3. pp. 249 to 255
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0601-0249$02.00/0

On Solutions that are Better Than Most

SANFORD BAUM

University of Utah

ROBERT CARLSON

Stanford University

(*Received October* 1978; *in revised form December* 1978)

A programmatic approach to solving decision problems by the use of 'better than most' (BTM) solutions is introduced. The BTM approach is contrasted to approaches that seek a solution whose outcome would be the best of all outcomes for feasible solutions, again according to criteria taken to represent the decision maker's objectives. This BTM approach is illustrated in terms of a decision problem in binary decision variables. This is followed by a discussion of (1) some of the probabilistic aspects of BTM solutions, including calculational experience, (2) a central BTM problem, akin to the problem of finding optimal methods in the more traditional approaches, (3) some *ad hoc* methods for overcoming this and other problems associated with the BTM approach, and (4) similarities and differences between the BTM approach and other approaches.

1. INTRODUCTION

ALMOST every normative method advocated for use in OR/MS problems seeks a 'best' solution, a solution whose outcome could be ranked over the outcomes of all other feasible solutions. Typically these methods successively examine candidate solutions and use the information obtained to generate new candidates. Thus, optimizing methods seek a 'best' solution by converging to the maximal (or minimal) value of a scalar function taken to represent the decision maker's goals. Within the last decade, proponents of multigoal approaches have called into question the desirability of representing all the decision maker's goal with a single function. In the multigoal approach, a vector function whose components represent the several goals of the decision maker is used, usually with the aid of dominance criteria. However, methods based on this approach are still directed towards searching for a 'best' solution as defined above. Even when it is recognized that a given normative approach will have to halt the search short of a 'best' solution (for example as a result of computer expense), the methods attempt to approximate a 'best' solution by using an optimal search routine that would achieve that solution given enough time.

It seems fair to say that the necessity of converging to a 'best' solution is taken to be *sine qua non* in the normative methods of OR/MS. Indeed to many, the inclusion of the word 'best' in the last sentence will appear to be either a tautology or redundancy. Such people will ask "If one does not search for the best, then for what does one search?" However, we take the position that in some instances, the price for avoiding the question of when a 'non-best' solution is good enough is too high. In most cases a 'best' solution for a normative formulation requires that the problem be mathematically tractable. This is another way of saying that the problem must be formulated in such a way that convergence to a 'best' solution is assured. Of necessity, (tractable) form becomes part of the problem, and the method becomes part of the solution. A 'best' solution to a problem formulated according to these procrustean requirements and subject to the imprecision accompanying communication on any complex question may be far from best in terms of the real problem.

Using the above viewpoint as a basis, we examine here an approach that does not require a 'best' solution. This approach looks for a solution whose corresponding outcome is 'good enough' or, more descriptively, is better than most. The approach is designed to be applied when an appropriate 'best' solution approach is lacking. The essence of the new approach is that while the problem formulation will continue to include the decision maker's goals and constraints, no attention will be given to formulating the problem in a way that will allow application of a search technique designed to converge to a 'best' solution. Instead, solutions will be randomly generated, examined for feasibility, and ranked according to the mathematical representation of the decision maker's goals. It will be seen that the outcome corresponding to a solution ranked highest of several hundred feasible solutions generated in this way is 'good enough' in the sense of being better than most. By this we mean that such a solution has a high probability of being ranked over a large fraction (e.g. 99%) of all possible feasible solutions, according to the mathematical representation of the decision maker's criteria.

We will illustrate this 'better than most' (BTM) approach with the aid of a problem in binary decision variables and discuss some of the probabilistic aspects of the BTM approach, including some relevant calculational experience. Next we will describe what appears to be the central BTM problem, akin to the problem of finding optimal methods in more conventional approaches, and suggest some *ad hoc* procedures for overcoming this and other problems associated with the BTM approach. We will conclude by discussing how the BTM approach relates to fuzzy concepts, simulation, and heuristic techniques.

2. AN ILLUSTRATIVE EXAMPLE

For purposes of illustration we have chosen a decision problem in binary variables. The problem may be visualized in terms of a situation in which a decision maker wishes to select projects in a way that will make use of his limited resources. His criteria for selection of projects involve two disparate goals which are combined into a single utility function. The assumption that the goals can be combined into a single function will be relaxed in subsequent discussion, so that the multigoal problem will be ultimately included in the discussion. To make a more explicit presentation and to relate the discussion to earlier work [1], we have chosen some previously considered problems where the goal and constraint functions are linear, and where the utility function is a Cobb–Douglas type. It will be seen that we can generalize from this specific problem and that the most significant assumption in the above formulation is that the decision variables are binary.

The mathematical form of the problem is as follows. We have a binary decision variable indicating rejection or acceptance of a specific project.

$$x_j = 0 \text{ or } 1 \quad j = 1, \ldots, m. \tag{1}$$

The two goal functions are given by

$$g_1 = c_0 - \sum_{j=1}^{m} c_{1j} x_j \tag{2}$$

and

$$g_2 = \sum_{j=1}^{m} c_{2j} x_j \tag{3}$$

where c_0 and the c_{ij} are positive numbers, since for mathematical convenience the goals will be required to be non-negative. In addition, selection of one or more projects uses a limited supply of n different resources, available in amounts b_i, i.e.

$$\sum_{j=1}^{m} a_{ij} x_j \leq b_i \quad i = 1, \ldots, n. \tag{4}$$

Finally, the decision maker's utility function is given by

$$U = g_1^f g_2^{1-f} \quad 0 < f < 1. \tag{5}$$

We emphasize that the outcomes corresponding to the various feasible solutions are well ordered. In the present case, this means that the function U can be used to order any outcome corresponding to a decision vector $\mathbf{X} = (x_1, \ldots, x_m)$.

When the utility function is known, this problem is really a single-goal problem which the conventional (optimizing) approach would formulate as a problem in optimization, i.e. to find a feasible decision vector $\mathbf{X}$, that will make the outcome, measured by the utility function described by (2), (3) and (5), as large or larger

Vol. 7
1979

than that for any other feasible vector. The BTM approach also seeks a feasible decision vector. However, the BTM approach merely requires that the decision vector eventually selected have an outcome that would have a high probability of having a utility function value greater than that of most other feasible decision vectors.

The next section will show that generation of a relatively few randomly selected feasible solutions will in fact produce a solution with the desired BTM characteristics. Furthermore, random generation of these feasible solutions is computationally simple when a sufficient fraction, say 1%, of the possible solutions are feasible, and when the decision variables are bounded. Consequently, the step from formulation to final solution with a BTM approach is a relatively simple one.

By contrast, the corresponding step in an optimizing approach is a formidable one. One method, based on the Frank–Wolfe algorithm [6] and employing the computer program for Hillier's 'bound and scan' concept [4] has been described [1]. However, the computer program is large, complex, and time consuming. Of course, other approaches (for example branch and bound) might be attempted; but success, in the form of a workable method developed in a relatively short time, is no longer guaranteed. Furthermore, inclusion of even one non-linear constraint relation, or possibly even a more complex objective function, would make success significantly more difficult to achieve, and might make it virtually impossible.

Our example can also be used to illustrate the use of the BTM approach in a multigoal problem. Assume, for example, that the decision maker's utility function (5) is not directly available to the analyst and that the only information he can obtain is the decision maker's reactions to a limited number of outcomes which are displayed to him. In this case the conventional formulation of the multigoal problem is to find a feasible decision vector, **X**, whose outcome, described by the goal function in (2) and (3), would be as much or more preferred than the outcome for any other feasible decision vector. However, the BTM approach merely requires that the feasible decision vector eventually selected have an outcome that would have a high probability of being ranked over most outcomes corresponding to other feasible decision vectors.

The next section will show that the generation of a few hundred randomly selected, feasible solutions will enable the analyst to interact with the decision maker and identify the solution whose outcome has the desired BTM characteristics. As in the case of a single goal, random generation of the relatively few feasible solutions is computationally simple. Thus the step from formulation to final solution in a BTM approach is also easy in the multigoal version of the single goal problem described earlier.

In contrast, the corresponding step in an approach that seeks a 'best' solution is even more formidable in this multigoal problem than in the optimizing case described earlier. The motivation for this paper springs from the conviction that when an analyst looking for a best solution is faced with this kind of problem, he will either avoid a programmatic approach entirely or formulate the problem so that a mathematically tractable method can be applied. In the latter case, the result is an optimal solution but only to a problem formulation that has been forced to fit the requirements of mathematical tractability. Given this kind of a result and the vagaries of human communication, one is no longer assured that a BTM solution to a problem formulated without the requirement for tractability is less valid than an optimal solution to the force-fit formulation. Clearly, there are many problems where optimal (and other 'best') methods are appropriate. However, there are a number of circumstances where a BTM approach should also be considered.

3. PROBABILISTIC ASPECTS OF THE BTM APPROACH

In this section we describe a BTM solution in quantitative terms. This definition is then applied to the question of how many feasible solutions must be examined in order to generate a BTM solution. The problem of how to treat multigoal problems where a clear-cut ranking criterion is not available is also discussed.

Recall that by a BTM solution we mean a feasible solution whose outcome would have a high probability of being ranked over most

of the outcomes for other feasible solutions. We assume that all the outcomes can be ranked on the basis of some explicit criteria, with ties being both possible and resolvable on some arbitrary basis. We define a 'high-ranked' solution to be one whose outcome belongs to the upper p per cent of the ranked outcomes. If we generate a feasible solution vector randomly, the probability that the corresponding outcome will satisfy our definition is p. If we randomly generate N feasible decision vectors, the probability that precisely Y of the corresponding outcomes will fall into the 'high-ranked' category has a binomial distribution with parameters N and p. Let p_0 be the probability that we do *not* obtain at least one such high-ranked solution (p_0 can also be described as the probability that $Y = 0$). Finally, suppose that we identify the solution whose outcome is ranked higher than the outcome for any of the other $(N - 1)$ randomly generated solutions. In terms of the original definition, this solution is a BTM solution if the probability of the solution being ranked in the upper p per cent is high. The complement of this statement is that the value of p_0 is low.

Assume, for the moment, that p is fixed, and recall that the parameters of the binomial distribution are p and N. We can then ask what are the values of N that will give us a low value for p_0. We have

$$p_0 = (1 - p)^N \tag{6}$$

from which we obtain

$$N = \log p_0 / \log(1 - p). \tag{7}$$

Figure 1 shows some typical values of N and p_0 for fixed values of p. This figure shows that generation of a few hundred feasible solutions will produce a result with a high probability of being ranked above 95% of the feasible solutions, generation of under 1000 will produce a result with a high probability of being ranked above 99% of the feasible solutions, and generation of under 10,000 will produce a result with a high probability of being ranked above 99.9% of the feasible solutions. Examination of (7) shows that for fixed values of p, a tripling of the value of N will reduce p_0 by a factor of 10^{-3}, which means, for example, that such a tripling will increase the probability of obtaining a BTM solution for a given p from 0.900 to 0.999.

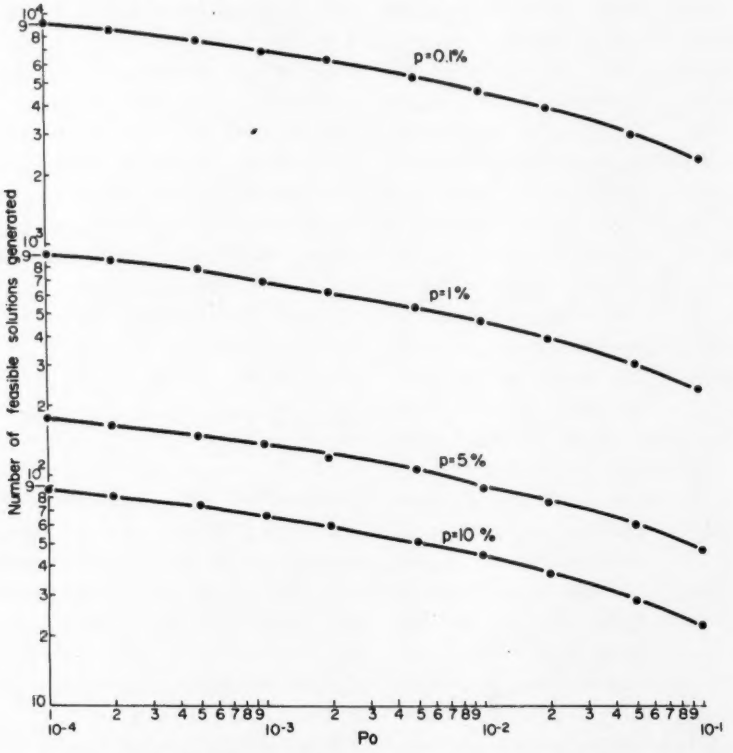


FIG. 1. *Number of random feasible solution required as a function of* p_0 *and* p *where* p_0 *is probability of not obtaining solution in the upper* p *percentile.*

We have now shown that random generation of only a relatively small number of feasible solutions will produce a BTM solution. Note that the only assumptions made were that (1) we could rank the outcomes from the randomly generated set, and (2) there were a large number of feasible solutions to draw from. We now turn to a brief discussion of these two assumptions.

For a simple problem in which the decision-maker's objectives are described by a scalar function of the decision variables, the procedure for ranking is straightforward. The value of this function is used to rank the outcomes. However in a multigoal problem, the decision-maker's objectives are described by a vector-valued function of the decision variables. More explicitly, the several goals, which are scalar functions of the decision variables, are components of a vector; and there is no explicit scalar function by which to rank the outcome vectors. In this case, it is necessary to identify the randomly generated decision vectors whose outcomes are undominated by the outcomes for other solution vectors from the randomly generated set. These undominated outcomes are displayed to the decision maker, who selects from among them the outcome he most prefers. The vector corresponding to this most preferred outcome is the BTM solution.

In making this last statement we have assumed that the decision-maker's preferences are 'rational' in the sense that he will always prefer outcome B to outcome A if B dominates A. Under this assumption the most highly ranked outcome in the randomly generated set must be contained in the subset of undominated outcomes from that set. Thus, the decision-maker's selection of his most preferred outcome from this subset identifies the BTM solution. A limited amount of experience with the generation of these kinds of subsets [1] indicates that they normally contain a small number of elements. Our experience also indicates that there are promising approaches to further reduce the number of outcomes that the decision maker must examine.

The second assumption for multigoal problems is that there are a large number of feasible solutions to draw our random sample from. This assumption implies that feasible solutions constitute at least a modest fraction, say 1%, of the total number of randomly generated solutions. When this condition holds, only a small amount of computer time may be required to randomly generate a set of feasible solutions. Our experience, based on solving a limited number of bicriterion problems in binary variables [1] leads us to believe that randomly generated solutions can be generated and processed at a rate of 1000 to 5000 per min, and that improvements in this rate may be possible.

4. TRACTABILITY IN THE BTM APPROACH

It can be seen from the preceding discussion that a BTM solution is characterized by the assumptions that (1) random generation of solutions is possible and (2) at least a modest fraction of the solutions so generated will be feasible. Generation of a BTM solution vector requires that a set which includes all feasible values (typically a range) be identified for each component and that elements from these sets be selected on a random basis. Ideally, these sets would contain only these elements that are found in feasible solutions, but this condition is not necessary. It is sufficient if all the feasible elements are embedded in a larger set.

Since the decision variables of the illustrative problem were binary, explicit identification of all possible set elements posed no difficulty, and problems with such decision variables appear to offer a fertile ground for BTM applications. However, explicit identification of all possible set elements is not necessary. Another attractive possibility for the BTM approach appears to lie in problems where the variables have explicit upper and lower bounds. But even this condition is not necessary. In many cases it will be possible to deduce the necessary bounds by examination of the problem constraints. For example, many problems contain one or more linear constraints and, if necessary, can usually be transformed into problems where all the decision variables are non-negative. If the linear constraints define a closed feasible region, then an upper bound for a given variable may be found by solving a linear programming problem whose objective function to be maximized is that variable and whose constraint equations are the linear constraints of the original problem. Lower bounds

may be found by finding the minimum value for the same LP problem. Of course, bounds derived in this fashion are not necessarily least upper (or greatest lower) bounds, since they do not reflect limits imposed by nonlinear constraints. But the values obtained with the LP process are bounds since the inclusion of other constraints can only restrict the feasible values further. More generally, and for the reason just noted, any constraint relations to which an optimization procedure may be applied can provide bounds for one or more variables. Thus there appear to be many problems where it is possible to identify ranges for the values of the decision variables.

A serious computational problem arises when the probability is small that a randomly generated solution is feasible. The cost of generating enough feasible solutions may be prohibitive. This situation is likely to occur either when the upper and lower bounds on the values of the elements of the decision vector encompass many non-feasible values or when the solution is tightly constrained (i.e. the set of feasible solutions is small relative to the number of solutions that could be randomly generated). Unless schemes can be developed, based on an examination of the constraints, to prevent the generation of many non-feasible solutions, the BTM approach is not well suited for the solution of these problems. It is fortunate, however, that these situations are precisely those in which implicit enumeration techniques, such as branch and bound, perform best. Many branches of the branch and bound tree can be pruned early because of the absence of feasible completions of partial solutions.

In summary, there appear to be a number of problems where it is relatively easy to randomly generate feasible solutions. When this is not true, use of a BTM approach may be a case of out of the frying pan and into the fire since one avoids the tractability problems associated with a procedure for finding a 'best' solution only to encounter another tractability problem in the random generation of feasible solutions. We have not yet developed criteria for recognizing such problems ahead of the event.

5. DISCUSSION

Casual questioning of a number of practicing decision makers who have taken OR courses has led us to the belief that although these practitioners value OR, they do not often use it to obtain answers to their decision problems. When questioned more closely, many of these people have said that OR concepts are useful for understanding problems but not for finding the solutions to them. It is also significant that most OR courses and texts place emphasis on programmatic techniques such as linear programming, whereas the most frequently applied OR techniques appear to be in such areas as descriptive PERT/CPM and simulation. Thus there are a number of indications of discrepancies between OR as taught and OR as applied. Part of these discrepancies may be attributed to the emphasis on finding the 'best' solution.

The BTM approach appears to offer an attractive alternative which, while maintaining the traditional problem formulation framework, avoids the conventional approach to finding an answer to such a problem. More specifically, the BTM approach utilizes decision variables, criteria for recognizing a desirable outcome, constraint relations and other general characteristics of mathematical modeling. At the same time the BTM approach avoids the difficulties that sometimes occur when formulating a problem so that a 'best' solution may be found. It is clear that there are many problems where these difficulties can be overcome, and it is equally clear that the BTM approach avoids these difficulties by requiring another form of tractability. However, we have already identified several problem types where the BTM approach does not encounter tractability problems. Consequently, we believe that the BTM approach merits further study.

Before concluding this discussion, we will make a few remarks on how the BTM approach relates to three other areas—'fuzzy' concepts, simulation, and heuristics. We begin with the interaction between the BTM approach and 'fuzzy' concepts, noting that the very phrase 'better than most' is a fuzzy phrase. The phrase is fuzzy in the same way that taller than most is a fuzzy phrase. Furthermore, a BTM solution is fuzzy in the sense that there are an infinite number of ranking values of confidence pairs (p, p_0) associated with any BTM solution. For example, the BTM solution obtained from the random generation of 100 feasible solutions will result in 99.4% confi-

Vol. 7
1979

dence of obtaining at least one outcome ranked in the upper 5%, and 86.8% confidence of obtaining at least one outcome, ranked in the upper 1%, etc. A BTM solution is also 'fuzzy' in the sense that, in a vast majority of the cases, the precise value of the outcome cannot be stated. Finally we may note that although Zadeh recently stated[1] that fuzzy concepts are descriptive rather than prescriptive, the fuzzy concepts applied here have led to a plausible prescriptive (or normative) approach. Thus the BTM approach may also be interesting as a prescriptive application of fuzzy concepts.

It may also be useful to distinguish the BTM approach from simulation techniques since they have features in common, and either can probably be regarded as an extreme form of the other. However, it is reasonable to say that in simulation both the selection of values for decision variables and the interpretation of the outcome resulting from the selection of those decision variables must be done externally. In this sense the simulation is descriptive rather than prescriptive, and it is this characteristic that distinguishes the BTM approach from simulation techniques.

Some of the work done under the rubric of heuristics, e.g. the recent work of Berry *et al.* [2] and of Mabert and Whybark [5], also involves the use of randomly generated solutions. However, the main thrust of such work appears to be the improvement of the probability of generating an improved solution, with the random generation process providing a benchmark against which other (non-random) generation methods are measured. However, to the best of our knowledge the approach described here is unique because it (1) offers an alternative to 'optimal' or other forms of 'best' solutions, (2) does not make any explicit attempt to find an improved solution, (3) includes multigoal problems, and (4) is based on some simple and primitive probability and ranking concepts.

In conclusion, we may state that the BTM approach offers a new and attractive point of departure in the normative analysis of decision problems. While the approach includes a number of useful characteristics from other approaches, the BTM approach has some unique characteristics of its own.

[1] Tutorial session at the ORSA/TIMS meeting in San Francisco, May 8–11, 1977.

REFERENCES

1. Baum S (1975) *Analysis for interactive solution of bicriterion problems in binary variables.* PhD Dissertation, Stanford University.
2. Berry WL, Marcus M & Williams G (1977) Inventory investment analysis using biased sampling techniques. *Mgmt Sci.* **23**, 1295–1306.
3. Henderson JM & Quandt RE (1972) *Microeconomic Theory, A Mathematical Approach.* McGraw-Hill.
4. Hillier FS (1969) A bound-and-scan algorithm for pure integer programming with general variables. *JORSA*, **17**, 638–679.
5. Mabert VA & Whybark DC (1977) Sampling as a solution methodology. *Decis. Sci.*, **8**, 167–179.
6. Zangwill WI (1969) *Nonlinear Programming. A Unified Approach.* Prentice-Hall, Englewood Cliffs, New Jersey.

ADDRESS FOR CORRESPONDENCE: *Professor Sanford Baum, Department of Mechanical and Industrial Engineering, MEB 3008, University of Utah, Salt Lake City, Utah 84112, USA.*

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 3. pp. 257 to 262
© Pergamon Press Ltd 1979. Printed in Great Britain

0306-0483/79/0601-0257$02.00/0

Linear Programming: Some Unsuccessful Applications

LEWIS J CORNER

University of Sussex

(*Received December* 1978)

Linear Programming is a widely used technique with many successful applications to its credit. However, for a variety of reasons it is found that LP often yields disappointing results. Three small industrial problems with obvious LP structure are described, and reasons for the lack of success of the technical approach are discussed. Some elementary but commonly ignored advice on methodology is suggested to potential LP users.

INTRODUCTION

THE LITERATURE of Operational Research includes a vast number of papers describing algorithmic developments and successful applications of mathematical programming. Less publicity is given to unsuccessful applications of Linear Programming, which is regrettable since one of the most important skills an OR analyst can acquire is the ability to recognise when a standard technique is likely to yield useful or disappointing results.

The business of practical OR is to solve problems in the best possible way. For this purpose an extensive array of techniques has been developed, but this paper attempts to show that the benefits of applying theory may be negligible for a variety of reasons.

It can be argued that we learn from our mistakes, and grow complacent from our successes. Consequently there should be a place in the literature for admissions of error, providing that such case studies describe pitfalls which other analysts would do well to remember and avoid.

This paper narrates three slightly fictionalised and disguised industrial assignments undertaken by the author some years ago as a junior practitioner. In each instance the problem under discussion had an obvious LP structure. In simplified form they all might have appeared in the LP section of an OR text or examination paper. Thus the methods of solution seemed straightforward and, by calling upon mathematical techniques, professionally satisfying.

However the outcomes were uniformly unsuccessful for reasons which should be apparent in the accounts of each project, and which are amplified in the conclusion.

The paper is primarily addressed to less experienced practitioners who are still in the process of establishing connections between their theoretical knowledge and perception of real problems. Older analysts will probably have similar memories to look back on, and may derive some nostalgic pleasure from these confessions of youthful misjudgement.

PROBLEM 1: ALLOCATION OF PRODUCTION RESOURCES

The main production department of a firm contained four machines, of different ages and speeds, which performed the same basic process. Several hundred products passed through the department in their first stage of manufacture. Set up times were negligible and standard speeds per batch were specified for each product on the machines by which it could be processed.

Allocation of batches to machines was done manually by a planning clerk who compiled

the weekly work programmes. The wide range of processing speeds between products and machines appeared to create a non trivial allocation problem. At the suggestion of the OR section it was agreed that the efficiency of the manual method of loading machines should be evaluated by using linear programming to determine the optimal solution for any given week.

An outline model was developed in about five minutes:

Suppose there are n products ($j = 1, \ldots, n$)
m machines ($i = 1, \ldots, m$)

and

t_{ij} = the batch time for product j on machine i
r_j = the number of batches required of product j
h_i = the number of hours available on machine i.

Let x_{ij} = the number of batches of product j produced on machine i. A suitable model is:

$$\text{minimise} \sum_i \sum_j t_{ij} x_{ij} \qquad (1)$$

subject to

$$\sum_i x_{ij} = r_j \quad (j = 1, \ldots, n) \qquad (2)$$

$$\sum_j t_{ij} x_{ij} \leqslant h_i \quad (i = 1, \ldots, m) \qquad (3)$$

all

$$x_{ij} \geqslant 0.$$

The objective function (1) seeks to minimise the total processing time and does not account for different machine costing rates. (2) constrains the output per product to the required number of batches. Constraints (3) ensure that no machine capacities are exceeded.

There was no restriction against splitting the batches of a product between two or more machines, and the numbers of batches involved were sufficiently large for a continuous solution to be acceptable.

A simple matrix generator was written, and the LP was run to discover the optimal plan for several actual recent weekly programmes. Machine availabilities were taken to be the loading implied by the manual schedule. Each LP included about 600 variables and 200 constraints.

The OR section was astonished to find that the LP results were virtually identical to the manual plan with respect to overall machine loadings. Had an uneducated genius been discovered who could solve quite large linear programs in his head?

The firm was reassured by the evidence confirming the efficiency of its seemingly primitive machine allocation methods, but before leaving the project the OR section decided to seek an explanation for the unexpected result.

It was found that batch speeds exhibited a structure which had not previously been investigated or suspected. The data had initially been tabulated in the form below:

	Machine			
Product	1	2	3	4
1	t_{11}	t_{21}	t_{31}	t_{41}
2	t_{12}	t_{22}	t_{32}	t_{42}
n	t_{1n}	t_{2n}	t_{3n}	t_{4n}

process times per batch (t_{ij}).

Closer inspection of the data revealed that the table could be presented in a different way:

	Machine			
Product	1	2	3	4
1	k_1t_1	k_2t_1	k_3t_1	k_4t_1
2	k_1t_2	k_2t_2	k_3t_2	k_4t_2
n	k_1t_n	k_2t_n	k_3t_n	k_4t_n

process times per batch ($t_{ij} = k_it_j$).

To demonstrate the significance of this structured table, consider a manually produced solution where all machines are fully utilised.

Suppose a batch of product i is switched from machine 1 to machine 2. k_1t_i hours are freed on machine 1 and machine 2 is overloaded by k_2t_i hours. This overload can be corrected by removing k_2t_i/k_2t_j batches of product j from machine 2. If these batches are now assigned to the underloaded machine 1, the time required on that machine is $k_1t_j(k_2t_i/k_2t_j) = k_1t_i$ hours. Thus the effect of making the exchange is to leave both machines fully loaded.

Clearly the same argument applies to all product/machine interchanges, hence for the LP model all feasible solutions have the same objective function value.

Although this problem gave every appearance of being ideally suited to a LP approach, the surprising conclusion was that any solution obeying the machine capacity constraints was bound to be optimal. The OR suggestion that an allocation problem existed was an illusion and the manual system of production planning was retained.

Vol. 7
1979

PROBLEM 2: CONTROL OF EXCESS INK STOCK

A firm produced a range of printed textiles using several hundred coloured inks, each print run required between one and six inks. The design of the printing machines caused the rollers to run dry when approximately ten litres of ink was still circulating in the troughs, reservoirs and connecting pipework. All inks were mixed and colour matched at the beginning of a run, and the average amount of each colour remaining when the order was complete was about 12 litres.

The normal practice was to send these 'left overs' to a mixed ink stockroom, and to draw from stock only when the exact colour was in demand. Since the interval between demands for any ink averaged several weeks, large stocks were carried. Quite often an ink was not easily accessible, having disappeared under a pile of 'kettles' in a rather badly organised storage area. Existing stocks were not infrequently overlooked, and the store was liable to contain several batches of the same colour.

New product ranges using incompatible but equally expensive inks were being planned. Anticipating an explosion of ready mixed ink stocks, the printing manager invited an OR analyst to consider the problem. The obvious approach, said the manager, was to devise a system of converting ready mixed inks into nearby colours to accelerate stock turnover and reduce stock levels.

All the coloured inks were mixed from a fairly small number of basic ingredients—pigments, solvents and thinners—and exact formulations by volume were defined for each of the many colours.

For any day the ink requirement and present stock position were known, and the basic ingredients could be assumed to be in unlimited supply. Obviously the problem of holding mixed ink stocks at the lowest possible level could be formulated as a linear program:

Suppose there are n inks $(j = 1, \ldots, n)$
m ingredients $(i = 1, \ldots, m)$

and

a_{ij} = number of litres of ingredient i per litre of ink j
S_j = number of litres of ink j in stock
R_j = number of litres of ink j required.

Let

x_{kj} = number of litres of ready mixed ink k to be used in making ink j
y_{ij} = number of litres of ingredient i to be used in making ink j.

A suitable model is:

$$\text{maximise} \sum_k \sum_j x_{kj} \tag{1}$$

or

$$\text{minimise} \sum_i \sum_j y_{ij} \tag{1'}$$

subject to

$$\sum_j x_{kj} \leqslant S_k \quad (k = 1, \ldots, n) \tag{2}$$

$$y_{ij} + \sum_k a_{ik} x_{kj} = a_{ij} R_j \quad (i = 1, \ldots, m;\ j = 1, \ldots, n) \tag{3}$$

all

$$x_{kj}, y_{ij} \geqslant 0.$$

The objective function maximises the volume of ready-mixed ink used in making the current requirement for fresh ink (1), or equivalently minimises the volume of basic ingredients used (1'). Constraints (2) ensure that ready mixed inks can only be used up to their stock levels. The last set of constraints (3) guarantee that the formulation of each ink made from ready mixed stocks and/or basic ingredients meets the specification.

With 500 colours the formulation is impressively large even by modern standards, involving some 250,000 variables and 3000 constraints. However the LP for a daily mixing programme could be very significantly reduced in size. Any colour including an ingredient not present in the formulation of another colour could not be converted to that colour, hence the corresponding variable need not be defined. Also, not more than about 20 colours would be needed on any one day.

A matrix generator would be needed to produce an efficient model for a particular day, but it was clear that the LP would be of manageable size—typically 400 variables and 200 rows.

Having proved the feasibility of a linear programming approach it was necessary to ask whether this provided a reasonable practical solution. Matrix generating and report writing programs would be needed, as would up to date stock records. The cost of designing and

implementing a suitable computer system, even with an immediately available LP package, seemed high in comparison with the likely savings. It was also apparent that daily printing production often differed from the published schedule, and frequent manual interventions would further reduce the potential benefits of using LP. After some discussion the OR analyst and the printing manager agreed that the technical attraction of using LP was more than offset by the cumbersome nature of the support systems that would be necessary.

Eventually a sub-optimal but more practical solution was adopted. For each ink a set of similar inks was defined, and tables prepared to show how a similar ink should be mixed with basic ingredients to produce the required ink. In the hands of the ink mixing foreman this proved to be an acceptable and efficient method.

Finally—and this perhaps should have been the opening question—the OR analyst asked if the 12 litres or so of excess for each colour when it was used was really unavoidable. The printing machines were general purpose printers designed for long runs and cheap inks. To avoid frequent topping up, the troughs from which the rollers picked up ink were of high capacity, far higher than was needed for the short printing runs which were usually made. By a simple replacement with smaller troughs it was possible to reduce the ink left at the end of a run to about 6 litres per colour, halving the initial problem at a stroke.

PROBLEM 3: PANEL CUTTING IN A FURNITURE FACTORY

As an exercise in product diversification a firm launched a range of self-assembly 'knock-down' furniture. Each item in the range comprised a set of rectangular laminated chipboard panels plus the associated fixtures and fittings. Panels were cut from larger laminated chipboard sheets, and concern was expressed about the proportion of this material which was being wasted in the form of unusable offcuts. The OR section was invited to consider the problem of reducing waste.

About 20 different panel sizes were required to produce the full range, and the economics of the laminating operation dictated that only one size of chipboard sheet could be used. Effectively the furniture factory was faced with a trim loss problem of intermediate difficulty, and LP was the obvious technique to use.

The only cutting restriction was that 'stopped' cuts were not possible, that is all cuts must be made across the full length or width of the board presented to the saw.

In theory LP models are easily constructed for such problems by associating a variable with every permitted cutting pattern indicating the frequency with which it is used. In practice the number of cutting patterns is generally too large to be handled, and the definitive work of Gomory and Gilmore suggests alternative LP based approaches for a variety of trim loss problems.

However for this problem the panels could be divided into two distinct groups, large panels and small panels. Only one large panel could be cut across the width of the basic chipboard sheet, producing an offcut which could then be used for the smaller panels. By assuming that the first cut would always be made lengthwise at one of five widths, the original chipboard sheet could be regarded as one of five combinations of a large panel sheet and a small panel sheet.

The number of sensible cutting patterns for large and small panels was by this means very significantly reduced, and an exhaustive list was produced manually in less than a day. The LP, which would have been extremely large and awkward to define for 20 arbitrarily sized panels, was shrunk to a model involving less than 150 variables as a consequence of the actual panel dimensions.

Full details of the panel sizes and the formulation are not included as they have little significance beyond the particular problem, but it is worth emphasising that a mixture of common sense and ingenuity will often allow a model to be constructed relatively easily even though the general type of problem under attack may be almost intractable.

Before OR advice had been requested cutting schedules had been prepared manually, making available adequate data for the comparison of LP and manual solutions. When the LP was run to determine the minimal waste for a given panel requirement it was found that the total percentage of waste could not be brought down by more than 2%. The maximum savings, excluding the cost of running the

Vol. 7
1979

LP every week, were estimated at less than £10,000 per year.

Further discussion showed that even this sum was largely illusory. No account had been taken in the model of the difficulty in using some of the patterns, especially the cost of adjusting the saw for different cuts and the possibility of mistakes if a large number of different cutting patterns were to be specified in any week.

The OR section were able to point out a few patterns which the manual planner had overlooked, and these alone would halve the difference between the manual and LP results if included in future cutting schedules.

Thus the true cost savings associated with using LP to minimise waste were newly estimated at not more than £3000 per annum, excluding the costs of implementing the necessary systems. The final benefits would be negligible, if they existed at all. It was agreed that the manual methods should be retained and that further OR advice should only be sought if there were any substantial changes to the production facilities or the product range.

In retrospect it was perhaps not surprising that LP had so little to contribute. As the analyst working on the problem had discovered, it was fairly straightforward to compile a list of efficient cutting patterns, and this was precisely what the manual planner had done. Given this list good solutions could be guaranteed, and manual methods could take into account production difficulties which were not easy to incorporate in the LP model.

Once again the OR section had found that a fairly small 'LP problem' could be solved with minimal loss of efficiency by a reasonably intelligent planner completely unversed in OR techniques.

CONCLUSION

Each of the three problems described provides a salutary example of an OR analyst attaching too much faith to the potential of linear programming. In this concluding section suggestions are made which a practitioner would be well advised to consider before embarking on any substantial LP exercise.

The most fundamental piece of advice is never to rush into a mathematical approach without mulling over the problem in a more informal way. In particular, wherever possible the physical system giving rise to the problem should be carefully studied. The printing ink study shows how a simple engineering modification, which had been overlooked by a departmental manager and his foremen, would immediately bring significant benefits. All that OR was doing was to point out that an apparent constraint could easily be changed. It is probably true that in any LP study greater gains will be achieved as a result of showing that apparently inviolable constraints can be relaxed than by optimising within the framework of currently accepted constraints.

Practical problems displaying LP structures are often very tightly constrained, allowing little scope for further improvement. The machine allocation problem is an extreme case where all feasible solutions are optimal, and perhaps the author was unlucky to discover such a problem so early in his career. A sensible approach, where it is suspected that gains may be small, is to perform a limited pilot study involving only the most significant variables. In the example cited a fraction of the products accounted for a high proportion of the total output. A preliminary LP study restricted to the most important products would rapidly have generated the results which ultimately emerged, and shown that a full scale study was unnecessary. An eccentric, but by no means foolish practice is to maximise and minimise the same objective function. It will often be found that the difference between the worst and best solutions is slight, and that the current policy has evolved, through trial and error, to something very close to the optimum.

In all three problems the analyst was in the fortunate position of having access to highly reliable data, and fairly simple models could be constructed. Generally life is less obliging, and there will be doubts concerning the accuracy of both the data and the model. Such models may indicate promising benefits which are never realised after a system has been implemented. Even worse, solutions may be accepted because of the benefits they predict, whereas actual outcomes lead to a deterioration in performance. There have been several unpublished examples in Britain of very large scale LP systems which have proved to be costly failures although they were initially implemented with a high degree of optimism. Whilst the literature abounds with examples of LP ap-

plications it is rare to find definitive statements about the net gains which were universally recognised as having been achieved.

Finally it must be realised that LP methods are by no means the only way of obtaining optimal solutions to all LP problems. This point is often obscured by the way LP is taught. Text books abound with exercises where solutions cannot be seen by inspection, but in practice intuitive solutions may be remarkably efficient. A good example is the classical travelling salesman problem. If the distances (or costs, times *etc.*) can be represented in the plane, the human brain is quite capable of seeing efficient circular tours where even the largest computers would run into difficulties.

The object of this paper is not to claim that all LP studies are misguided or unsuccessful; such a suggestion is manifestly absurd. However it is the author's feeling that the benefits of LP have been oversold. The techniques are attractive because of their simplicity and the multiplicity of problems to which they can be applied. It is impossible to be competent at OR without appreciating the battery of mathematical techniques which now exist, but equally it is impossible to be competent without a sound judgement of when they should be used.

JK Galbraith in *The Affluent Society* coined the phrase 'conventional wisdom' to describe a framework of economic belief which he regarded as untenable despite its widespread popularity. The OR analyst must of course always suspect the conventional wisdom of the organisation where he works, for without this challenge there will be no change. On the other hand he must also realise that often, whether by accident or design, the conventional wisdom is precisely what it claims to be, and that no amount of mathematical modelling will invalidate the current solutions to problems he is invited to study.

ADDRESS FOR CORRESPONDENCE: *Lewis J. Corner, Esq., Bel Air, 2A Newfield Road, Newhaven, Sussex.*

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci. Vol. 7. No. 3. pp. 263 to 268
© Pergamon Press Ltd 1979. Printed in Great Britain

0306-0483/79/0601-0263$02.00/0

International Register of Theses and Dissertations

Wood JR (1978) The Client–Consultant Relationship—Optimizing to Meet the Needs of Corporate Management, with particular reference to the Industrial Marketing Research Function. MBA Thesis, University of Alberta. (Available at: University of Alberta, Edmonton, Alberta, Canada T6G 2M7)

The following are obtainable at the MIT Microreproduction Laboratory, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02139, USA:

Abramson FD (1977) Strategic Planning Systems in Large and Small Organizations. SM Thesis, MIT, Cambridge, MA.

Alexander M (1976) The Property Tax in Cambridge: the Effect of One Hundred Percent Evaluation. SM Thesis, MIT, Cambridge, MA.

Alexander ML (1975) Canadian Commercial Bank Strategies Towards International Expansion. SM Thesis, MIT, Cambridge, MA.

Al-Bassam NI (1977) Integration of Corporate Long Term Planning and Specialized Functional Planning. SM Thesis, MIT, Cambridge, MA.

Amano Y (1976) A Study of Corporate Growth. SM Thesis, MIT, Cambridge, MA.

Anderson RD (1975) Forest Resource Dynamics. SM Thesis, MIT, Cambridge, MA.

Anderson S (1975) Attitudes Toward Physical Working Spaces and Their Implications for Managers. SM Thesis, MIT, Cambridge, MA.

Andreoli PN & Steedman J (1975) Management Decision Support Systems: the Impact on the Decision Process. SM Thesis, MIT, Cambridge, MA.

Arnt-Jensen TJ (1976) Key Variables—a Comparative Study in the Oil Industry. SM Thesis, MIT, Cambridge, MA.

Arita S (1977) The Application of MBO in Planning and Control Systems for Japanese Corporations. SM Thesis, MIT, Cambridge, MA.

Assad AA (1976) Solution Techniques for the Multi-Commodity Flow Problem. SM Thesis, MIT, Cambridge, MA.

Aurich LD (1976) A Computer Code for Solving Integer Programming Problems with Variable Resource Levels. SM Thesis, MIT, Cambridge, MA.

Avallon JC & Stolper D (1975) Marketing Strategy and Product Line Planning for a Small Business Firm. SM Thesis, MIT, Cambridge, MA.

Baggeroer WL & Fox J (1975) Centralized, Decentralized, or Distributed: a Preliminary Model for Computer System Configuration. SM Thesis, MIT, Cambridge, MA.

Bajoria GG (1975) Transfer of Technology: Problems in Adaption of Foreign Technology in LDC's. SM Thesis, MIT, Cambridge, MA.

Bannister TC (1977) Structured Thesis on Adaptive and Integrative Planning. SM Thesis, MIT, Cambridge, MA.

Barata De Arujo F (1975) Monetary Correction of Fixed Assets and Working Capital. SM Thesis, MIT, Cambridge, MA.

Basile PS (1975) A Socio-Economic Model of the US—Evaluation of Transition Policies. SM Thesis, MIT, Cambridge, MA.

Bautista AJ & Dill D (1975) A Rational Method for Lease Analysis. SM Thesis, MIT, Cambridge, MA.

Beall AO Jr (1976) Dynamics of Petroleum Industry Investment in the North Sea. SM Thesis, MIT, Cambridge, MA.

Bell MA (1976) Evaluation of the Management Control System of Alcan Jamaica Limited. SM Thesis, MIT, Cambridge, MA.

Bellafato FJ & Nanfeldt C (1975) Automobile Workers—How Dissatisfied Are They? SM Thesis, MIT, Cambridge, MA.

Berger HC (1976) The Comprehensive Employment and Training Act: The Massachusetts Balance of State Prime Sponsor. SM Thesis, MIT, Cambridge, MA.

Bhavnani RH (1976) Performance of Direct and Portfolio Foreign Investments in the Absence of Technological Advantages: the Case of the Food Processing Industry. SM Thesis, MIT, Cambridge, MA.

BRACKEEN RE (1975) Spending Oil Wealth: a Study of Iran's Strategies for Allocating Oil Revenues to National Development and Foreign Policy Goals. SM Thesis, MIT, Cambridge, MA.

BRUCE MR & HUNT DW (1976) Communications in Non-Union Companies. SM Thesis, MIT, Cambridge, MA.

BUTKOVICH GJ (1977) Financial Management of Oil Pipelines. SM Thesis, MIT, Cambridge, MA.

CALLAHAN PR (1977) A Study of Barriers to Upward Mobility for Women in a Manufacturing Facility. SM Thesis, MIT, Cambridge, MA.

CAMPBELL MJE (1976) A Functional Model for a Mail-Order Marketing Division. SM Thesis, MIT, Cambridge, MA.

CARROLL DJ Jr (1977) Quality of Work Life, Is the Supervisor Ready? SM Thesis, MIT, Cambridge, MA.

CHEN EF (1976) Foreign Exchange Exposure Management: a Linear Programming Approach. SM Thesis, MIT, Cambridge, MA.

CHERTKOW L (1976) The Effects of Capitalization of Long-Term Leases on the Retail Industry. SM Thesis, MIT, Cambridge, MA.

CLARK JP (1975) An Economic Analysis of Substitution Between Copper and Aluminium in the United States. SM Thesis, MIT, Cambridge, MA.

COMBS J (1977) Electric Automobiles: Industry and Government Issues. SM Thesis, MIT, Cambridge, MA.

CONDIT PM (1975) Financial Analysis of Capital Budgeting in a Mature Airline Industry. SM Thesis, MIT, Cambridge, MA.

CONNORS PF (1976) The United States Fishing Industry. SM Thesis, MIT, Cambridge, MA.

COX GT (1976) Trends in Commercial Aircraft Marketing. SM Thesis, MIT, Cambridge, MA.

CRAWFORD JE (1976) An Evaluation of Some Impacts of Collective Bargaining on Productivity. SM Thesis, MIT, Cambridge, MA.

CUMMINGS W (1975) Two-World Dynamics. SM Thesis, MIT, Cambridge, MA.

DANIELS JE & TSIANG R-H (1975) The Marketing of a Business Journal: a Strategy for the *Sloan Management Review*. SM Thesis, MIT, Cambridge, MA.

DE WECK P (1976) Changes in the Nuclear Power Programs of France, West Germany, Sweden and Switzerland Between 1973 and 1975 Compared to the US Experience. SM Thesis, MIT, Cambridge, MA.

DEMAIN C (1976) A Determination of the Strengths and Weaknesses of Long-Range Planning. SM Thesis, MIT, Cambridge, MA.

DIENER A (1975) Regional Shopping Center Development: the Evaluation of a Proposed Site. SM Thesis, MIT, Cambridge, MA.

ECKBO PL (1976) The International Petroleum Market—Some Behavioral Options. PhD Thesis, MIT, Cambridge, MA.

ENSER L (1977) Operations Management: a Case Study. SM Thesis, MIT, Cambridge, MA.

ERICKSON DC (1976) The Use of Simulation for Project Risk Analysis. SM Thesis, MIT, Cambridge, MA.

ESSER AM (1976) Minority Shareholders in Close Corporations. SM Thesis, MIT, Cambridge, MA.

ESSER K (1976) Hedging Against Inflation Risk in Domestic and International Corporate Finance. SM Thesis, MIT, Cambridge, MA.

EVANS PAL (1974) The Price of Success: Accommodation to Conflicting Needs in Managerial Careers. PhD Thesis, MIT, Cambridge, MA.

EVANS RL & SMITH GT (1976) A Study of Management Planning, Control, and Information Systems in Thirty Medical Schools. SM Thesis, MIT, Cambridge, MA.

FAMA F Jr (1975) Labor Turnover in the Rubber Industry: a Case Study at the Plant Level. SM Thesis, MIT, Cambridge, MA.

FAN PY (1976) Computational Problems in Modeling the Oil and Gas Discovery Process. SM Thesis, MIT, Cambridge, MA.

FARBER MA (1976) Making Group Decisions with Limited Resources. SM Thesis, MIT, Cambridge, MA.

FAY J-M (1976) A Framework for Computer-Based Cost Evaluation of Highway Projects in Developing Countries. SM Thesis, MIT, Cambridge, MA.

FENNER PD (1975) Participative Management: a Lesson in the Real World. SM Thesis, MIT, Cambridge, MA.

FIEBELKORN SL (1976) Management Control in the Decentralized Corporation. SM Thesis, MIT, Cambridge, MA.

FILUSCH EJ (1975) The Effects of the Asset Depreciation Range System of Accelerated Depreciation on Corporate Income Taxes, Cash Flows, Market Value and Cost of Capital. SM Thesis, MIT, Cambridge, MA.

FISHER I (1975) Towards Estimating the Dependence Structure of International Security Returns. SM Thesis, MIT, Cambridge, MA.

FOX HI (1976) Going Private. SM Thesis, MIT, Cambridge, MA.

FRENCH SJ (1977) An Investigation into the Nature of the Demand for Passenger Air Travel. SM Thesis, MIT, Cambridge, MA.

Frenkel W & Jalan M (1975) Design and Evaluation of a Model for International Funds Management. SM Thesis, MIT, Cambridge, MA.

Friedman AN, Stefanini M & Witherell R (1975) The Impact of the Sloan Fellowship Program on Sloan Fellow Graduates from 1968 to 1974. SM Thesis, MIT, Cambridge, MA.

Friesen GF (1976) The Case Against New Restrictions on Foreign Direct Investment in the United States. SM Thesis, MIT, Cambridge, MA.

Garman RB (1977) Management Planning and Control Systems for the *Sloan Management Review*: a Case Study. SM Thesis, MIT, Cambridge, MA.

Garrett WA Jr (1976) Temporary Management Structures: Strategies for the Implementation of Interstate Telephone Network Engineering Projects. SM Thesis, MIT, Cambridge, MA.

Gates AB & Bobo CJ (1977) A Look at Risk Attitude via a Project Management Game. SM Thesis, MIT, Cambridge, MA.

Geschwind JK (1975) Work Force Innovation: a Chrysler–UAW Four-Day Work Week Experiment. SM Thesis, MIT, Cambridge, MA.

Getchell EH (1975) Factors Affecting Employee Loyalty. SM Thesis, MIT, Cambridge, MA.

Gianoukos PC (1975) A Simulation Model of a Suburban Community. SM Thesis, MIT, Cambridge, MA.

Gilliam AF (1976) The Employee Retirement and Income Security Act of 1974 and the Construction Industry: the Limitations to Governmental Regulation. SM Thesis, MIT, Cambridge, MA.

Goehrke GT, Walthall WW & Kaplan DJ (1977) The Deterrent Effect of Capital Punishment: a Statistical Model. SM Thesis, MIT, Cambridge, MA.

Goode BB (1976) US Public Policy Regarding Domestic Satellite Communications. SM Thesis, MIT, Cambridge, MA.

Goodie LS (1977) Planning the Installation of a Data Base Management System. SM Thesis, MIT, Cambridge, MA.

Gorman K (1975) An Application of Dynamic Programming to Inventory Control in the Publishing Industry. SM Thesis, MIT, Cambridge, MA.

Gordon EQ (1977) Attitudes toward Work and Satisfaction with Job Security. SM Thesis, MIT, Cambridge, MA.

Greer PS (1975) Segmentation Issues in Marketing an HMO. SM Thesis, MIT, Cambridge, MA.

Guggenheim AH (1977) Strategic Planning in a Multinational Company: a Case Study. SM Thesis, MIT, Cambridge, MA.

Haas EA (1977) Strategies of the Automobile Industry in an Increasingly Energy Conscious Environment: the Effect of the Energy and Conservation Act of 1975. SM Thesis, MIT, Cambridge, MA.

Hague DM (1977) A Comparative Study of the Induction of Automobile and Computer Technology. SM Thesis, MIT, Cambridge, MA.

Hamilton JT (1976) Productivity Improvements in Health Care Through the Utilization of High Technology Product Developments. SM Thesis, MIT, Cambridge, MA.

Harbur MM (1977) Adaptation and Integration in the Planning Process: a Case Study Using the Lorange Framework. SM Thesis, MIT, Cambridge, MA.

Haug PR (1976) Technology Transfer in Government Bureaucracies. SM Thesis, MIT, Cambridge, MA.

Hays GG (1977) An Analysis of Factors Influencing Industrial Participation in the MIT–Industry Polymer Processing Program. SM Thesis, MIT, Cambridge, MA.

Heard ML, Jackson RJ & Smith CD (1976) Analysis of Engineering Retention Programs and Their Impact on Organizational Dynamics. SM Thesis, MIT, Cambridge, MA.

Helmrath TA (1977) Planning and Control in Academic Medical Centers: a Case Study. SM Thesis, MIT, Cambridge, MA.

Helton SL (1977) A Capital Budgeting Case Study. SM Thesis, MIT, Cambridge, MA.

Henry AB (1976) Construction of Market Correlated Portfolios. SM Thesis, MIT, Cambridge, MA.

Henry PA (1977) Investment in European Production Facilities by US and UK Pharmaceutical Firms. SM Thesis, MIT, Cambridge, MA.

Herman GA (1975) What is Success? A Survey of Placement Personnel. SM Thesis, MIT, Cambridge, MA.

Higginbottom M & Madsen J (1975) The Building Construction Industry: a Search for Improved Quality Control. SM Thesis, MIT, Cambridge, MA.

Hill RM (1977) An Analysis of Airline Profitability. SM Thesis, MIT, Cambridge, MA.

Ho YFF (1977) An Analysis of the Characteristics and Needs of the End-Users of the Data Processing Services. SM Thesis, MIT, Cambridge, MA.

Hogan MG (1976) Capital Budgeting in the Offshore Oil Industry. SM Thesis, MIT, Cambridge, MA.

Holcomb CC (1975) A Survey of Attitudes of Engineers about Management of Corporations. SM Thesis, MIT, Cambridge, MA.

HOLLENBERG T (1977) Risk-Sharing in the Establishment of a Private Sector Uranium Enrichment Industry. SM Thesis, MIT, Cambridge, MA.

HOLLINGS PF & RAFF G (1975) The Effects of Debt Indexation of the Value of the Firm. SM Thesis, MIT, Cambridge, MA.

HOOD TJ (1976) A Study of the Determinants of Borrowing Costs for Issuers of General Obligation Municipal Bonds. SM Thesis, MIT, Cambridge, MA.

HOROWITZ RJ (1976) Consulting for a Small Publishing Company. SM Thesis, MIT, Cambridge, MA.

HOVELL CR (1977) An Investigation of Line-Staff Relationships in the Implementation of Financial Planning Models. SM Thesis, MIT, Cambridge, MA.

HSIUNG D-I (1977) Some New Perceptions of Achievement Orientation among Young Men and Women. SM Thesis, MIT, Cambridge, MA.

HUGUENIN M (1975) Wage Determination in the Public Sector: a Case Study. SM Thesis, MIT, Cambridge, MA.

HVEREN HU (1975) Economics of North Sea Crude Oil Ventures. SM Thesis, MIT, Cambridge, MA.

JAIN BC (1976) Development of Business Strategy in an R & D Organization—a Case Study. SM Thesis, MIT, Cambridge, MA.

JAMERSON JE (1977) Bankruptcy Costs: a Further Examination. SM Thesis, MIT, Cambridge, MA.

JEAN DF (1977) Inflation and Financial Reporting Systems: a Case Approach. SM Thesis, MIT, Cambridge, MA.

JEFFERSON TG (1975) Human Resources Management: a System Dynamics Study. SM Thesis, MIT, Cambridge, MA.

JOHNS JH (1975) Intelligent Computer-Aided Dispatching for Urban Police Patrol Units. SM Thesis, MIT, Cambridge, MA.

JOHNSON EC Jr (1975) Engineers: Interdisciplinary Non-Homogeneity and Job Satisfaction. SM Thesis, MIT, Cambridge, MA.

JONES RC (1977) Using System Dynamics in Business Policy Planning. SM Thesis, MIT, Cambridge, MA.

KAFKA BD (1977) The Multinational Corporation and Political Risk. SM Thesis, MIT, Cambridge, MA.

KANODIA AS (1976) Material Requirements Planning—a Study. SM Thesis, MIT, Cambridge, MA.

KAPLAN N (1976) Architectural Profiles: Notes on the Socialization of Architects. SM Thesis, MIT, Cambridge, MA.

KAVANAGH GL (1977) The United States Shipbuilding Industry and Influences of Conglomerates. SM Thesis, MIT, Cambridge, MA.

KESRI M (1977) An Examination of the Transfer of Technology between Developed and Developing Countries. SM Thesis, MIT, Cambridge, MA.

KHANNA A (1975) Developing Management Training in India. SM Thesis, MIT, Cambridge, MA.

KHILNANI A (1975) An External Resource Allocation Model for Less Developed Countries. SM Thesis, MIT, Cambridge, MA.

KIDDER NK (1975) The Dynamics of Revenue Sharing. SM Thesis, MIT, Cambridge, MA.

KIRK GD & HEDGES J (1975) An Assessment of Current Attitudes of Petroleum Industry Engineers. SM Thesis, MIT, Cambridge, MA.

KISSEL EW (1975) Attitudes of Supervisors Working Rotating Shift Schedules. SM Thesis, MIT, Cambridge, MA.

KITANO T (1977) Structural Problems of the Japanese Textile Industry. SM Thesis, MIT, Cambridge, MA.

KLEPPA PA (1977) Locations of Increased Oil Refining Capacity: the Issues, the Reasons, the Implications. SM Thesis, MIT, Cambridge, MA.

KLIMM RL & MONTGOMERY J (1975) A Study of the Application of Computer Technology in Retail Stores. SM Thesis, MIT, Cambridge, MA.

KOBAYASHI H (1977) Debt Repayment Capacity Analysis of Developing Countries: a Case Study. SM Thesis, MIT, Cambridge, MA.

KROLIKOWSKI MA (1977) An Assessment of Planning Needs and Capabilities of a Small Company. SM Thesis, MIT, Cambridge, MA.

KURTZ EW (1977) Hospital Information Systems: a Guide for Managerial Planning. SM Thesis, MIT, Cambridge, MA.

LAIDERMAN RA (1977) The Finances of Labor Union Locals and the Need for Financial Intermediation. SM Thesis, MIT, Cambridge, MA.

LARSON JH (1977) Offshore Liquefied Natural Gas Terminals. SM Thesis, MIT, Cambridge, MA.

LAURENT D (1975) Clustering Analysis Upon the Paris Stock Exchange. SM Thesis, MIT, Cambridge, MA.

LAURENT G (1976) A Dynamic Analysis of the Housing Market in Paris. SM Thesis, MIT, Cambridge, MA.

LEVINE ME (1977) Hierarchical Segmentation of the Margarine Market. SM Thesis, MIT, Cambridge, MA.

Linderman JL (1975) Time Pattern Analysis in School Scheduling. PhD Thesis, MIT, Cambridge, MA.

Lindsay CC (1976) Works Councils in the Federal Republic of Germany. SM Thesis, MIT, Cambridge, MA.

Livengood TC (1975) A Consideration of the Influence of Money Supply on Stock Market Prices. SM Thesis, MIT, Cambridge, MA.

Lloyd KR (1977) An Econometric Model of Extensions of Oil and Natural Gas Reserves in the United States. SM Thesis, MIT, Cambridge, MA.

Lotz SH (1975) The Role of the Private Sector in On-The-Job Training Training for the Disadvantaged. SM Thesis, MIT, Cambridge, MA.

Mandelbaum SF & Knopf R (1976) Success in the Mail Order Business—An Examination of the the Industry and an Analysis of Key Variables. SM Thesis, MIT, Cambridge, MA.

Mannheimer TS (1977) The Interrelationship of Career and Family: an Attitudinal Study of Young Management Professionals. SM Thesis, MIT, Cambridge, MA.

May LZ (1976) Capital Budgeting in Bell System Telephone Companies. SM Thesis, MIT, Cambridge, MA.

Mermelstein SE (1976) Factors Affecting Adoption of Solar Water Heating Systems. SM Thesis, MIT, Cambridge, MA.

Miller JR (1975) The Black Manager in Commercial Banking: a Case Study, SM Thesis, MIT, Cambridge, MA.

Mintz RD (1975) An Analysis of Collective Bargaining Structure in the Federal Government. SM Thesis, MIT, Cambridge, MA.

Murray JT (1975) Corporate Strategy for the 1980's: a Comparison of the World's Major Motor Vehicle Manufacturers. SM Thesis, MIT, Cambridge, MA.

Mathan LH (1977) Supermarket Stockouts and Brand Switching. SM Thesis, MIT, Cambridge, MA.

Nelson MH (1977) Mathematical Ability, Inability and Disability. SM Thesis, MIT, Cambridge, MA.

Nitay B & Zurr Z (1976) Computerization in the Newspaper Industry. SM Thesis, MIT, Cambridge, MA.

Okada M (1976) Diversification of Cigarette Manufacturers in the United States. SM Thesis, MIT, Cambridge, MA.

Pizano C (1976) Design of a Control System by Responsibility Centers for a Manufacturing Company in a Developing Country: a Case Study. SM Thesis, MIT, Cambridge, MA.

Powell Y (1976) A Systems Analysis of a Manpower Management Information System. SM Thesis, MIT, Cambridge, MA.

Riese PJ (1977) Formal Planning Systems: a Critical Evaluation of the State of the Art. SM Thesis, MIT, Cambridge, MA.

Rowan RE (1977) An Evaluation of the Use of the Multi-Regional Input–Output Model in Projecting State Industry Employment. SM Thesis, MIT, Cambridge, MA.

Runkle DL (1976) Planning: a Comparative Analysis. SM Thesis, MIT, Cambridge, MA.

Russell RC (1976) The Family of Operating Systems—a Test Case. SM Thesis, MIT, Cambridge, MA.

Schneckenburger WF (1976) Iran and Iraq: a Comparative Study of Economic Development in Petroleum Rich Countries. SM Thesis, MIT, Cambridge, MA.

Schnieder RJ (1976) Real Estate Investment Trusts: an Analysis of Recent Developments. SM Thesis, MIT, Cambridge, MA.

Scotch DS (1977) Medical Practice Plans: Strategic Business Units? SM Thesis, Cambridge, MA.

Shanley MR & Weston DW (1976) The Budgetary Decision Making System in a University Setting. SM Thesis, MIT, Cambridge, MA.

Shapiro WE (1975) Financial Accounting in an Era of Inflation. SM Thesis, MIT, Cambridge, MA.

Silverstein JE (1975) Economic and Political Factors in the Internationalization Strategies of Small Country Firms: the Case of Israel. SM Thesis, MIT, Cambridge, MA.

Soos JE (1975) Some Effects of Teachers' Unions on Public School Management. SM Thesis, MIT, Cambridge, MA.

Spiro HJ (1976) Problem Solving Approach to Air Pollution Control Administration. SM Thesis, MIT, Cambridge, MA.

Sprecher LT (1976) Residential Housing as an Investment: Its Rate of Return and Risk. SM Thesis, MIT, Cambridge, MA.

Stephens SB (1975) Containing the Costs of Hospital Care. SM Thesis, MIT, Cambridge, MA.

Stone TM (1976) The Impact of Inflation in Accounting on Rate of Return Regulation of Public Utilities. SM Thesis, MIT, Cambridge, MA.

Suzuki K (1975) Multinationalization of Steel. SM Thesis, MIT, Cambridge, MA.

Tani K (1976) Diversification in Japanese Firms. SM Thesis, MIT, Cambridge, MA.

Tomkins FC Jr (1976) Toward Optimal Utility in the Small Company Retirement Benefits Program. SM Thesis, MIT, Cambridge, MA.

TORIELLO LN (1977) Macrocephalous: a Regional Model of Urbanization Process in the Third World. SM Thesis, MIT, Cambridge, MA.

USHIYAMA M (1975) Measurements of Public Attitudes Toward Donating Blood. SM Thesis, MIT, Cambridge, MA.

VAN DEELEN MD (1975) Computer Simulation of a Business Firm. SM Thesis, MIT, Cambridge, MA.

WALLMAN SM (1977) The Police: an Organizational Approach. SM Thesis, MIT, Cambridge, MA.

WEIL JI (1977) A Study on the Nature of Small Shipments. SM Thesis, MIT, Cambridge, MA.

WEISS R (1977) The Financial Sector in the French Economy since 1946. SM Thesis, MIT, Cambridge, MA.

WICK CW (1975) The Management Side of Ministry. SM Thesis, MIT, Cambridge, MA.

WIELAND JW (1976) The Effects of Inflation on the Value of Firms. SM Thesis, MIT, Cambridge, MA.

WILSON LL (1977) Quaker Decision Making: Adapting the Friends' Method for Use in Non Quaker Organizations. SM Thesis, MIT, Cambridge, MA.

WINEMAN JJ (1975) The Energy–Environment Tradeoff: Impact of the National Environment Policy Act on Public Participation in the Nuclear Power Plant Siting Process. SM Thesis, MIT, Cambridge, MA.

WINN MJ (1977) A Study of the Factors which Determine the Success of a Small Business in the Convenience Store Industry. SM Thesis, MIT, Cambridge, MA.

XAFA A (1975) Switzerland: Financial Structure and Capital Flows. SM Thesis, MIT, Cambridge, MA.

YAMAGATA M (1975) Early Warning System of Bankruptcy. SM Thesis, MIT, Cambridge, MA.

ZEBRAK M (1975) The Coffee Cartel. SM Thesis, MIT, Cambridge, MA.

ZWEIG M (1975) 1975 Sugar Situation: Increased Production Needed. SM Thesis, MIT, Cambridge, MA.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

Vol. 7
1979

ol. 7
979

ERRATA

Omega, Volume 7, Number 4, 1979

The journal title in reference 23 on p. 307 should read:

Acad. Mgmt Proc. 33rd *An. Meet.*

The 7th line on p. 357 should read:

...the disks shown in these slides was superior to that...

In FIG. 2 on p. 365 the top of the left-hand axis should read: 140.

BLANK
PAGE

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 4, pp. 269–272
Pergamon Press Ltd 1979. Printed in Great Britain

0306-0483/79/0801-0269$02.00/0

EDITORIAL

Power Corrupts

GONE ARE the days when the manager could expect discipline, loyalty and good time-keeping from his subordinates while he would roll in at about ten, deal with the post, pronounce a verdict on some outstanding matter, go out for a three hour lunch, and be ready to leave the office at four in the afternoon for a round of golf with an important client. Also gone perhaps is the old premise that by being envious of the manager's life style, his subordinates became strongly motivated to work hard in order to attain a managerial status and enjoy similar privileges.

Nowadays it is difficult to find managers who do not work long hours, who do not worry, who are not candidates for an early ulcer or even a heart attack. If anything, the working life of the successful executive should serve as a deterrent to aspiring young managers to beware of the dire consequences of success. And yet, there are plenty of aspirants who strive and persist, and who are in a great hurry to get on, and when they do, they exert themselves to get further, or at least to protect their position from conniving understudies.

What is it that motivates these people? What is the reward of this relentless struggle to get to the top and to stay there? It certainly cannot always be money; there are many senior executives who have long satisfied all their material needs, and even in countries where harsh tax systems culminate in significant diminishing returns of net income, they prove to be a disincentive only to some but not to all. It has therefore often been observed that managers' ambitions extend to more than just financial rewards; they look for an intellectual challenge, for achievement, for status and recognition, and above all—we are confidently told—they look for *power*.

"Power implies that either the organization or the individual has the ability to coerce someone into doing something he does not want to do" [3, p. 105], or to put it more simply, power is "the ability to take action to effect a change in one's environment or in someone else's environment" [1, p. 4]. Power to make choices between alternatives or to influence a decision, power to impose or relieve constraints, power to specify procedures, power to affect behaviour through the specification of performance criteria—in short, power is an attribute

that embodies the aspiration of the individual to control his own destiny and the well-being of the system of which he is a part, and for such an individual the attainment of power and the ability to increase it are clear manifestations of success and a cause for great personal satisfaction.

Given, then, that power is something that many managers wish to have, the question arises as to how they go about attaining it. Observation suggests that some individuals devote a fair bit of their time to scheming and intrigue, that they plan confrontations and skirmishes with great attention to detail, and that their skill at crafty manipulation has been developed to a fine art. How did they acquire these skills? Have they learnt them just through experience and observation of other manipulators at work, or have they embarked on an elaborate and systematic strategem to master this craft?

I have always been naive enough to assume that most managers belong to the former group, that with their climb up the hierarchical ladder and the passage of time they become increasingly aware of the role of power and the machinations that it generates, and that whatever skills they acquire in this respect they do so almost imperceptibly through their own personal experience. Many people, of course, like to read about the Machiavellian exploits of great historical figures or captains of industry, and many delight in Stephen Potter's amusing principles of gamesmanship [2], but I have never thought that any one took such writings seriously as recipes for the acquisition of power. Perhaps Potter Academies for One-upmanship do in fact exist, but I confess that I have not come across them.

Vol. 7
1979

Sooner or later, though, someone was bound to take the subject seriously. I am not referring to research into the development and implications of power centres in industrial and social organizations (such research has been going on for a long time and is well documented in the literature), but to an attempt to provide a do-it-yourself manual for those in avid search of power.

Such a long awaited volume has now appeared, 'The Craft of Power' by RGH Siu [4], and as it is published by the well known publishing house John Wiley and Sons, my first reaction was that it deserved close scrutiny. On the dust cover the publisher confidently proclaims: "Personal power—acquiring it, maintaining and expanding it—has undoubtedly obsessed human beings since the first ape-men learned to stand erect. And those who seek power have never lacked for guides—sagacious or imprudent, ethical or amoral—from the *I Ching* to Machiavelli's *The Prince*, from the Bible to the Watergate Tapes." And he goes on: "Stressing self-discipline, the lessons to be learned from the masters of personal power, and the very *feel* of power and status, this is a deeply engaging, *transforming* book. An astute observation of human behaviour, a provocative revelation of the way in which

much of society is driven, *The Craft of Power* merits your attention. Read it today."

So I did, and a more pretentious pompous and repulsive document I have not seen for a long time. The very thought of painstakingly instructing people how to manipulate others for the sake of personal greed, I find positively abhorrent, and I was therefore looking for some redeeming feature that would justify the allocation of space to this volume on one's bookshelf. Alas, I found none. I imagine that the author sat down one day and said to himself, "I have to find a suitable topic that has such a wide appeal in society, that cuts across the confines of particular professions, and that is relevant to all walks of life, and that topic would then be a best-seller". But it does not take a great deal of browsing in the book to realise that he has failed miserably.

Leaving aside, for the moment, the ethical justification of a guide of this kind, the only saving grace of the book is its list of contents, which vaguely conveys a promise of some interesting topics to look forward to, but the text is utterly lacking in depth and the monologues are more reminiscent of a sales pitch than a serious attempt at the development of a set of postulates or arguments. The one thing that the book excels in is a series of contrived aphorisms, often cloaked in bewildering obscurantism. Here are some priceless gems from the sections entitled 'manœuvering and striking' and 'negotiating and pressing on' [4, pp. 21–26]:

"Coil yourself omnidirectionally"
"Intensify the hopelessness of the exploited masses to resist. Mix blame with pain in your punishments. Tailor the compulsions"
"Prevail through adaptive centrality"
"Remain unpredictable. Preserve surprise. Ambiguity, inconsistency, evasion, aphonia. Conceal the cunning through deception of diversion"
"Pry psychologically with the crowbar of time"
"An innings ends, but the game of power goes on. The confrontation being seeded overlaps the negotiation being matured. Do not strike out by breaking the continuum."

I refuse to go on. The quotations which summarize the author's thinking speak for themselves, and the subsequent chapters that purport to elaborate on these themes are not much better. What I find utterly baffling is that a publisher of such high repute as Wiley could have rated the book as a potential winner, to the extent that the blurb on the dust cover ends with the promise that the author's "next book will be devoted to executive serenity". The thought makes your blood curdle. It also makes you wonder whether people should be allowed to use valuable timber on such questionable projects.

But let me return to the much more serious issue which all professional people have to address themselves to, namely whether they need concern themselves with the purpose for which their clients seek

advice. Many would argue that the high moral tone implied in the comments made on Sui's book are misplaced, and that as a professional you have no business to set yourself up as a judge of your client's motives. If he asks you to improve his cash flow, his profitability, his chances to beat his opponents, the saleability of his products of services—you readily oblige. Your advice is normally not conditional on your approval of his moral character, of the goods that he sells or of the use that he proposes to make of his profit. Similarly, on a personal level, many professionals help clients to improve their appearance, looks, physical condition, speech. Why should you treat power differently, you may ask?

I take the view that the purpose for which advice is solicited *is* relevant, that if the client is evil-minded and intends to use the advice he is given for evil ends, then you should not be instrumental in helping him to achieve his purpose. And as for the conscious unadulterated lust for power, I am suspicious of the motives involved and of the use to which it might be put. But many will profoundly disagree with this point of view and persuasively argue for keeping craft and ethics apart.

Vol. 7
1979

SAMUEL EILON
Chief Editor

REFERENCES

1. EILON S (1979) *Management Control*. Pergamon Press, Oxford, UK.
2. POTTER S (1947) *The Theory and Practice of Gamesmanship*. Rupert Hart-Davis, London, UK.
3. SCANLAN B & KEYS JB (1979) *Management and Organizational Behavior*. John Wiley, New York, USA.
4. SIU RGH (1979) *The Craft of Power*. John Wiley, New York, USA.

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 4, pp. 273 to 274
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0801-0273$02.00/0

FEEDBACK

Planning, Futures, and the Skeptics

'LONG-RANGE planning is a fraud'; 'No one can predict the future'; 'Planning is worthless when everything changes so rapidly'; 'it is the unexpected events that are determining'—so say the skeptics, in increasing numbers, about the role of planning and futures in modern society. Most of these criticisms betray some basic misconceptions about the meaning of futures and planning. Of the many definitions of planning, Peter Drucker's captures its essence best: "Planning deals with the futurity of present decisions." Futures provides the long-term perspective for linking planning and decision-making. By separating our perceptions of the future from our decisions about the present, skeptics perpetuate a know-nothing and do-nothing mentality.

The future is unpredictable in an absolute sense. This is true, whether we are trying to predict events in the physical world, such as earthquakes and droughts, or in the social and political world, such as revolutions or assassinations. Of course, greater predictability and certainty is possible when our state of knowledge about system behavior is greater or when the possibility of intervention from outside the system being examined is smaller. Compare, for example, the relative predictability of eclipses versus the predictability of world climate. An eclipse of the sun could be altered by a sizable physical body entering our solar system. Similarly, forecasting world climate changes becomes considerably more difficult when human intervention—either intended (e.g. deliberate climate modification) or unintended (e.g. release of carbon dioxide from fossil fuels)—is at play. So it is largely a matter of *degree* of predictability and forecastability—acknowledging that at some point differences in degree may become, for practical purposes, differences in kind.

No one could have predicted Sadat's visit to Israel, including, very likely, Sadat himself. So we should not try. However, any serious analyses of the possible future course of the Israeli/Arab conflict would have almost certainly included as an alternative the possibility of a settlement along the lines now being pursued. True, the probability assigned to such an alternative would have been low. Nevertheless, the value of considering a range of possible alternatives—even somewhat improbable ones—is to force oneself to think carefully about how to confront such eventualities.

Thus, the basic purpose of planning and futures cannot be to predict but rather to provide guidance on how we can better cope with uncertainty, the constant condition of humankind. This uncertainty stems from a number of causes: incomplete understanding, unexpected events, individual choice. And these will always be with us. Statements about some future state of the social, political, and even the physical world should invariably include, explicitly or implicitly, room for alternative outcomes. Demographers, for example, should be forced to provide a *set* of at least three forecasts of US population; economists should be required to tell us about the *probable range* of inflation or unemployment; climatologists should describe a range of likely world climate *scenarios*; and so forth. There are many reasons why this is not done. At times, forecasters—particularly in the government sector—are put in positions of stating what someone may *want* to happen rather than what the forecaster *thinks* will happen. Acknowledgement of uncertainty may be equated in the minds of some to imprecision or evasion (an 'intellectual cop-out'). Or forecasters may succumb to the natural temptations of playing the role of clairvoyant, speaking out as if they understand a situation better than they actually do.

Even if forecasters behave sensibly, unexpected events will still come along, perhaps with increasing frequency, as the world becomes more complex and intertwined. But life is not determined solely by the unexpected. If it were, then our strivings to achieve understanding and some measure of control of our collective destiny would indeed be absurd. Rather, we are in a race with complexity; we need to think through a range of futures resulting from the interplay of expected and unexpected events so that we may decide how we can increase, even if only slightly, the probability of more desirable futures.

How can this best be done? In a variety of ways. But here I want to focus on the role of the much maligned 'expert'. No matter how well informed and perceptive, experts cannot predict, so we should not ask them to do so. But this does not mean that we want to deprive ourselves of the most valuable resources we have, namely, the collective wisdom of our best intellects. What we need to learn to do is to use our so-called experts differently.

We need much better ways of defining and identifying those with insight and expertise. We need to produce some balance of undisciplinary and multi-disciplinary skills. We need to seek out individuals who are able to see beyond the narrow boundaries of particular disciplines and who have a flair for perceiving new possibilities, new opportunities, and new alternatives. Most importantly, we need to direct the attention of such individuals into telling us something about the nature of the 'game' we may be 'playing' and something about the odds or the range of outcomes. Paul Samuelson, MIT's Nobel-prize economist, has put this fairly succinctly in a recent interview:

"if someone asks me what the terms of trade between rural Iowa and urban Massachusetts will be in 1984 or 1994, I can tell him how to think about the problem but I cannot indefinitely narrow down the probabilities. The essence of a science is to know not just your strengths but also your weaknesses. Modern economics cannot prophesy accurately."

In essence, what Samuelson is saying is that we should ask our experts to shed light on the most important factors that may determine outcomes and how these factors appear to be related.

Let me illustrate how this has been done in a particular case. Few physical phenomena are as poorly understood as climate change and the underlying forces creating it. Yet during the past year, it has been possible for the first time to portray our collective state of knowledge (and ignorance) on likely world climate change to the year 2000. This has been done by carefully selecting a deliberately diverse group of some of the world's leading climatologists, eliciting their perceptions about particular key climate variables, such as global and regional temperature and precipitation, allowing each participant to self-rate himself and peer-rate others with respect to degree of expertise, surfacing (rather than suppressing) the degree of uncertainty associated with each response, and combining the aggregate judgments in the form of alternative climate senarios.

The perception of the 'expert' is not the only, or even the major, resource at our disposal. Not at all. We depend most heavily on the subjective judgment of individuals when we are unable to rely on history (because of little relevant data or the strong suspicion that history will not repeat itself) and when we do not know the important variables or do not understand the important relationships among them. If we can use prior history as a point of departure for deriving perceptions about the future, usually for short-term or well understood phenomena, then we can use the impressive array of statistical and correlative techniques available to the technician and analyst to detect underlying trends. Or if we are fortunate enough to have sufficient historical data and an adequate theoretical base, usually for physical systems, then we may be able to make predictions with a considerable degree of confidence.

But availability of adequate data and theory are not the usual conditions in social or political arenas. Here, greater emphasis must necessarily be placed on human (including expert) judgment working with scraps of data and hints of underlying theories. However, the essential steps in the process are the same, whether practiced by the physical scientist, the corporate strategic planner, or the political or social analyst: describe a range of possible *alternative futures*; assign (to the extent one can at first) *probabilities of occurrence* to each such future; define the principal *choices* available for influencing outcomes; evaluate the range of probable *consequences* stemming from each choice; select courses of *action* based on some chosen set of preferences; and—most important of all—*monitor* carefully the whole process of planning and choosing so that the process can be gradually improved as it is iterated over and over again. It is, in fact, the minute interweaving of planning and acting that is the essence of management of any kind—public or private, organizational or individual.

Viewing planning and futures as something distinct and apart from the practical world of decision-making and acting has perhaps been our most profound error. The separation is a reflection, in part, of our modern-day obsession to specialize and to carry reductionism to excessive lengths. But, in part, it is also due to a basic misunderstanding of the relationship of planning and futures to decision-making. Having broken apart the planning–acting 'Humpty-Dumpty,' it may be difficult to put him back together again. In the meantime, countless meaningless planning exercises take place, both in the private and public sectors, exercises in which the planning function has been emasculated by its separation from operational decisions. Today in many corporations, planning managers find themselves collecting, organizing, and analyzing information about possible changes in the corporate environment that line managers do not know how to use. And line managers go through pointless 'planning drills' with little notion of how such processes might in fact help their earnings performance next year, if not next quarter. Similar situations exist in many public sector activities where, for example, in the defense establishment, it is not uncommon to find those formulating strategic options for the late '80s working in splendid isolation from those who are making choices about such options.

Every decision or act makes a statement about the future. Every useful plan tells the decision-maker something about the nature and uncertainty of the environment that may confront him now and in the future. Effective decision-making and planning are thus as inextricably linked as opposite sides of the same coin. It is the failure to understand this intimate relationship that is proving costly to us all both in the misapplication of societal resources and in the failure to realize achievable societal goals.

ROY AMARA

The President
Institute for the Future
2740 *Sand Hill Road*
Menlo Park
CA 94025
USA

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 4, pp. 275 to 286
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0801-0275$02.00/0

Change, Clarity and Clausewitz

DUNCAN DAVIES

Chief Scientist and Engineer, Department of Industry

(*Received January* 1979)

The framework of laws which in the past has governed the interaction of technology and business is proving inadequate as the scale and interconnectedness of economic, social and technological problems accelerate. The author evinces the need for new systems of control wherein both the macro and micro levels of perception of and response to large scale, multifaceted problems can be harmonised.

1. INTRODUCTION

a. *The subject*

THE OBJECT of integrating disciplines and subjects which do not share models, perceptions, or even vocabularies is obviously needed in real life, but because of the diversity and conflict of models it is difficult to secure accord and harmony when one speaks of method and process. What is usually beneficial under these circumstances is to try to achieve mutual understanding in terms of general aims—to agree about ends, and to use this agreement in reconciling different approaches to the means. Scientists and engineers in industry and the public service are accustomed to the pursuit first of their own subjects and projects, and then of economic growth, and competitive success and survival, through technological change and development. Sociologists and economists are concerned about the impacts of these processes and ambitions on people and organisations, and with such matters as full employment, stable international exchange, and the fruitful exercise of comparative advantage. Can we assist the process of unification by addressing ourselves to the simple question 'what for?'

b. *The analogy*

I propose to seek guidance from an analogy that at first may seem slightly bizarre, and which certainly must not be pushed too far. This is the comparison of technological change with pre-1900 war. Both have been the source of social disturbance and reorganisation. Both have involved discomforts and suffering for many affected by the process; both, however, have been pursued because of their fruits and hoped-for outcome. But above all, both initially have been the subject of muddles and very confused thinking, again because of the unrecognised differences between the models of different groups of people. For the engineer, technical change is the means for plying his trade, as war was for the General. For the embattled environmentalist, nuclear reprocessing or fast breeders may represent an unmitigated evil, as war was and is for the pacifist. For the politician, some of his voters want supersonic airliners because of jobs or views about progress, others oppose them because of noise, cost, or waste, just as voters have been ambivalent about war, although I have to agree that technological change has hitherto been a positive-sum-game in which gains have massively outweighed losses, while war has often been a negative-sum-game.

My reason for pushing ahead with the analogy despite this major difference is that war, its purposes, and its processes have been dispassionately and competently studied synoptically whilst technical change has not. The impact of the classical work, 'Vom Kriege' written

[1] This paper is based on the Joan Woodward Memorial Lecture given by the author on 28th November, 1978.

by General Clausewitz in 1830, has continued ever since. It coldly analysed the position, motivation, constraints and objectives of aggressors and defenders, haves and have-nots, and includes the effects of warm emotion and hot anger. He answered the question 'what for?' by going further than the simple concept 'to win', for he recognised the potential impermanence of victory; the result of war, he says, is never absolute. "The conquered state sees in it only a passing evil, which may be repaired in after times by means of political combinations". And, dare I add, bringing matters up-to-date, by export policies based on highly competent, selective, and single-minded technical change. By way of experimental evidence, Clausewitz was able to quote from the campaigns of generals throughout history, but particularly from those of Napoleon, whose synoptic view gave such a striking advantage. On this basis, he began the whole work: "We propose to consider first the single elements of our subject, then each branch or part, and, last of all, the whole, in all its relations—therefore to advance from the simple to the complex. But it is necessary for us to commence with a glance at the nature of the whole, because it is particularly necessary that in the consideration of any of the parts their relation to the whole should be kept constantly in view".

He then distilled his views into the words: "War is therefore an act of violence intended to compel our opponents to fulfil our will". This escapes from the unsatisfactory statement that the purpose is 'to win' or, more emotionally, 'to protect civilisation' 'to defend the Fatherland', 'to look after our women and children', which confuse and divide those who wish to think about war as an element of policy. Clausewitz' statement can be used by a pacifist, for whom the words 'an act of violence' may carry the corollary 'unacceptable'. It applies equally to aggressor and defender, and requires no judgements as to provocation or historical precedent. And it carries the clear implication that policies pursued in war must in the ensuing peace be pursued by other means if the fruits of victory are to continue.

In this context, it is interesting to extend the terminology of conflict as has been done by Rapoport. He defined the object of a *fight* as being the destruction of an opponent, of a *game* as that of outwitting him, and of a *debate* as that of convincing him. Hence, bearing in mind considerations of revenge following defeat in war and resentment following defeat in games, transfer into debate and consensus emerges an essential need for permanent policy.

Thus, Clausewitz, who took fourteen years to write his uncompleted book (the most important part of which is in sketch form, for he died, aged 51, before any of it was published) established a pragmatic model regarding war as part of politics. This view profoundly affected the history of the nineteenth and early twentieth century, for it permitted escape from the confusion of idealistic or cataclysmic views, that helped to start and prolong wars. Once modern weapons came in, acts of violence, though intended to compel opponents to fulfil the victor's will, did not do so. By its own showing the period of validity of 'Vom Kriege' came to an end. The same may well happen in the end to the analysis of technological change that we are now recommending, but a 100 years' validity would be distinctly useful.

Vol. 7
1979

c. *The confusion*

Present thinking about the objectives of technological change is as unsatisfactory as pre-Clausewitz thinking about war. Some say that technology is aimed at economic growth, but others then retort that this is temporary, since conventional economic growth, being dependent on finite resources, cannot be sustained for ever. Others speak of 'better living', and are then met with observations that one man's meat is another man's poison. Some see technology as a basic vehicle for human ingenuity and competitiveness, justified on that account. One of the most interesting analyses is that of the late Professor Fred Hirsch, who wrote of the conflicts within and between what he called the paradox of affluence, the distributional compulsion, and the reluctant collectivism. The first of these states one of our most important current technological perplexities:

'Why has economic advance become and remained so compelling a goal to all of us as individuals, even though it yields disappointing fruits when most, if not all of us, achieve it?'

One of the prime causes of muddle is the sheer diversity of the impacts of technology,

and the insufficiency of the means for generalising them so as to provide some kind of objective function that can then be optimised. We have a polynomial of the form

$$\text{social effect} = f(x_1) + f(x_2) + f(x_3) \ldots f(x)$$

where

x_1 = extension of life expectancy
x_2 = mobilisation of new energy sources
x_3 = effect on climate of CO_2 from furnaces etc.

perhaps for a further hundred terms and where there are plenty of cross-terms.

No one has yet been able to follow Clausewitz by defining even an approximate model that will adequately bring together the Friends of the Earth and the mining companies or nuclear energy agencies, or Ralph Nader and industry at large. It is fairly clear that I am not going to produce such an analysis in this paper.

The next step is to assign priorities to the major impacts so as to reduce the problem to manageable dimensions. For the criterion to be continually applied must be 'no matter how long the interval over which we are seeking to plan, what do we do about it all next Monday'. A look at historical and present trends may enable us to approach this assembly of a hierarchy of priorities, about which much discussion will be needed.

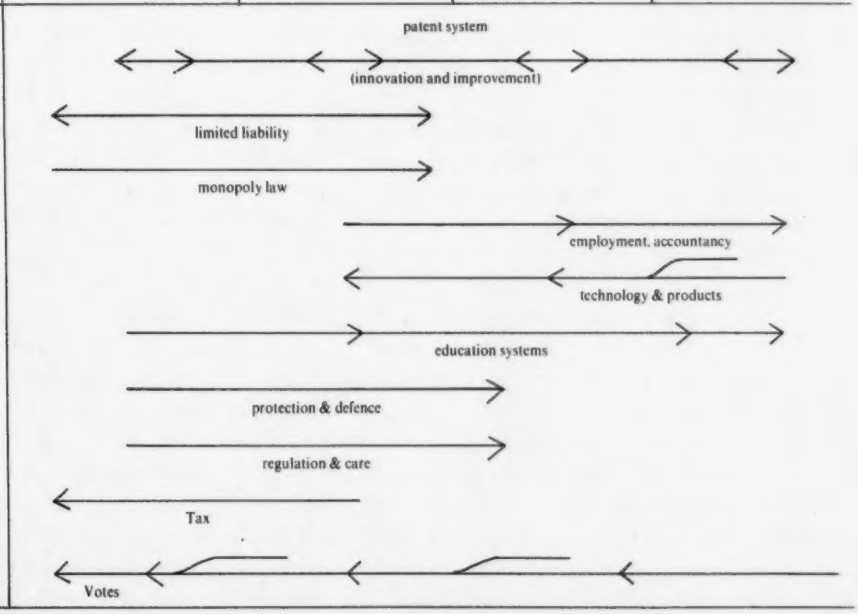
2. THE INTERACTIONS BETWEEN SCIENTISTS, ENGINEERS, BUSINESSMEN, AND POLITICIANS

a. 1650–1950

Over the period following the Renaissance, it was found advantageous (partly because of the 'avoidance of bureaucracy') if different social groups could enjoy freedom from detailed mutual interference. Each needed the outputs of the other groups, but felt that they were best without detailed advice. CP Snow spoke of two cultures: one could easily speak of a score or a hundred. It will suffice for our present purpose to speak of 4, some of whose interactions are described in the following model (Table 1).

Businessmen who defined their enterprises and set them up appropriately, are protected by the procedures of limited liability. They, and individual engineers, scientists or other inven-

TABLE 1. PRESENT SYSTEMS FOR BRINGING TOGETHER THE ACTIONS OF VARIOUS SOCIAL GROUPS.

Group	Politicians & state servants	Businessmen	Engineers	Scientists
Objectives	Acceptable & sustainable society, laws, & process	Comparative advantage (national, company)	Useful and reliable systems & equipment	(a) Natural understanding (b) Its application
Mechanisms for interaction and co-operation				

tors, are encouraged to disclose technology by the award to them, in return, of a temporary (*ca* 16 year) monopoly in the form of a patent: this encourages the diffusion and extension of any invention. Development of excessive power through monopoly is attacked by separate monopoly laws. (Factory legislation and environmental regulations do go into some detail). But the general methods have proved to be effective and durable for a long time.

The point in time at which the state began to make direct choices of technologies (or to take steps to encourage others to do so) varied between countries. Similar direct concern of businessmen with their detailed technologies also began at different times. In the UK, early technological pre-eminence underpinned the winning and development of an empire as a source of cheap raw materials and food. Thereafter, the country and its industries benefited most from the smooth and improving operation of this supply system, for which many of the most able were educated—hence the value attached to the humanities including politics and economics, which provided the right background for good colonial and home administrators. Engineering and science were applied with vigour in the Navy, but after the initial great industrial successes in the late eighteenth and early nineteenth centuries, technological initiatives often passed to other countries, notably US and Germany. USA, being short of people, built on standardisation and mechanisation processes in which businessmen became deeply and personally knowledgeable. Germany needed to add value cheaply, to offset UK raw material price advantages: hence her dedicated development of efficient production machinery. In both countries it was and is usual for company boards to contain many engineers proud to describe themselves as such. It was natural and necessary for engineers to get to know about foreign markets (for product design, maintenance, and adaptation) and about costs and profitability. The governments helped educationally by setting up Land Grant Colleges in USA and Technischen Hochschulen in Germany. Engineers were involved in the technical, production, marketing and finance policies of their employers. This was also true in France at the top where the Polytechniciens (and others from the Grandes Ecoles) brought together technology, economics, and politics both in companies and government, thus effectively uniting the strengths of the public and private sectors.

In Britain over the late nineteenth, and first half of the twentieth century, there was considerably less sharing of skills between scientists and engineers on the one hand and businessmen, politicians and top civil servants on the other. Scientists and engineers placed more reliance on innovation and quite new products, developed with fairly modest funds (by general industrial investment standards) in laboratories, workshops and small back rooms. Production technology was often (but not always) relegated to second place.

b. 1950–1980

The more recent awakening of interest of British engineers and scientists in economic and social factors, and of more businessmen and public servants in technical details, has arisen from three causes. The first is the sheer increase in importance of the technical component of innovation decisions. Entrance fees for major changes have escalated in real terms, largely because of the excellence, acceptance, and economy of scale and consequent low cost realised for the existing products.

Consequently, a decision to go ahead with a new design of aeroengine or model of automobile, or a quite new fibre, plastic or pharmaceutical, requires the commitment of sums—perhaps over a billion pounds—that are substantial in relation to the total funds commanded by a large company. To champion such an innovation, a scientist or engineer must understand the financial position, and to decide between this kind of enterprise and a less risky but less profitable expansion of present production, the businessmen must know more than ever about the detail of the new invention, its likely cost, its acceptability in the market, and the hazards.

The second cause is the increase in environmental impact, arising because of increasing scale of production coupled with higher standards for clean air and rivers. Technological changes and expansion are now challenged fiercely, both rationally and irrationally, so that businessmen, engineers, politicians, and public servants must be able to marshal and assess the arguments on both sides, and then conduct their case appropriately. Scientists

Vol. 7
1979

need new understanding in order to investigate the risks in the appropriate terms. It is interesting that engineers in Germany regard us in Britain as being fortunately rational in our approach.

The third cause is the great increase in the relative importance of production engineering and technology. Japan has shown that it is possible to prosper by excellent engineering that changes the production costs, reliability, quality, and design of existing types of product, without major technological invention related to the product. A good case in point is that of the Japanese motorcycle, which swept the world by brilliant design, production, and marketing, based partly on the observation that motorcycles now need bigger and more efficient electrical systems, including starters. By contrast, Britain has maintained its rate of invention, but repeatedly found that development without first-class follow-up in production technology simply presents competitors with a major opportunity. Computers and numerically controlled machine tools afford examples of inventions made and first developed in Britain, and then taken over and exploited by others through superior competence in quantity production and appropriate design.

c. 1980 *onwards*: *can technology be abandoned*?

The changes of which we have so far spoken are essentially defensive measures for British technology and business to prosper: the two cultures have needed to be introduced. To begin with, many did not like the change, for it can alter the balance of power in favour of those who can think and act on a broad base, including technology and microeconomics, and is awkward both for the classics-trained administrator and the specialist scientist. There is still strenuous opposition to educational programmes that sacrifice some specialist depth to breadth that includes technology and social studies. And although the opinion and influence behind the new need is now strengthening, it is now clear that we have to go further still. We now need an increasing concern with a total process of social advance that gets rid of the waste and frustration of adversary procedure. When the businessmen and engineers are united, they must then get down to a considerable measure of joint planning and design with those concerned about the state and its practices, for otherwise there are real dangers that the adversary methods will simply bring the whole process of technological change to a dead stop. This kind of argument is already being seriously conducted in USA, where robust concepts of democratic procedure regard too much collusion between regulators and regulated, or any other major estates of the union, as being suspect and the possible source of cosy arrangements and the over-concentration of power.

At this point there are voices that ask 'does it matter if we stage a moratorium on technical change while we pause for thought?' Alongside are other, more strident, voices that already assert that enough is enough, and that technology—especially if big—stands condemned. 'Back to the hills, forests, and fields tilled without industrial inputs' goes the cry: 'it was better in the past; small is beautiful'.

The most basic arguments concern food from usable land, available energy, and sustainable expectations of comfort, esteem, and longevity. First, regarding land, there are about ten billion acres (out of the world land total of 33 billion) that might be mobilised for agriculture. (At present, there are between 3 and 4 billion arable acres.) Thus, if world population levels out at 10 billion or thereabouts, there is about one usable acre per head. Secondly, regarding available energy, present consumption averages rather less than 2 kW per head, with a peak in US at 11 kW, and substantial numbers of people consuming between 0 and 2 kW. For everyone in a world of 10 billion people to achieve present American consumption levels, world energy production would have to rise from the present figure of 7.5 terawatts to over ten times that figure—perhaps 100. And we must note that American (and Western) aspirations are to increase their *per capita* energy consumption yet further.

There may be adequate methods for reaching these very high energy generation figures while living from one acre of usable land per head, but it seems much more likely that ingenuity will increasingly direct itself toward the achievement of satisfaction with less waste, so that we do not have to visualise 100 terawatts or more. There will be powerful arguments for such policies—some arising from tho[illegible]ts

about the environmental impact (atmospheric CO_2 levels and greenhouse effects; radioactive waste and other ash; appreciable climatic effects from heat dissipation). Others arise from the energy investment (and associated cash investment) needed to raise energy production to these levels. Similarly, there is scope for the application of extra ingenuity to the production of an adequate and varied diet from the one acre average. For a vegetarian diet, there is no problem at all, but substantial use of animal protein (whose production needs between 5 and 20 times as much vegetable protein) makes matters much tighter.

To satisfy social and material aspirations, it therefore seems that there will be a growing need for ingenious husbandry and substitution, and progressively less scope for resource—intensive economic growth with the current levels of waste. Such ingenuity will range over a wide area—more economical materials, with some recycle that does not need much energy; careful and varied food preparation, with less wasteful use of meat; methods of communication that rely less on personal movement. Perhaps most important of all could be the assignment of social esteem to ingenious effectiveness in getting more from less. The overall effect would seem likely to be an increase in the need for technology, but with a larger degree of built-in social design. From the viewpoints both of adequacy of output and of social acceptability, a 'primitive', non-technological approach looks very inadequate.

This being the case, how do we move on from the present hesitant and mutually defensive interaction of social and technological activities, to a more enthusiastic pooling of skills and showing of strengths? There is a historical precedent: Rousseau indicated that, in the late eighteenth century, man ceased to be involved in a social contract over whose shape he had no control: he was beginning to design the contract himself, by the democratic process. So it may be for the technological contract: technology was born free but is increasingly in chains. The important thing is for the technologists themselves to join in the forging of the social chains, so as to minimise the necessary constraints and accept them loyally.

3. THE TECHNOLOGICAL CLAUSEWITZ

a. *The objective and the method of approach*

Having asserted that the continuation of technology is necessary, and urged that this proposition be supported and clarified by as much evidence as is possible, we are able to give a very simple, and pragmatic Clausewitz-style answer to the question 'what is technical change for?' in the words 'to survive'. But we then have to sub-divide, by considering the time-scale and the different groups to which we belong. Table 2 does this: it presents a *possible* view of a hierarchy of needs for three groups of which we are members, the smallest group (UK) having its own need to add to those of the two larger groups 'Western Democracies' and 'Human race'. We could, of course, add further groups moving upward, such as 'mammals' and 'living systems' and having entries here such as 'preserve the ozone layer' and 'limit hard radiation levels'. But we have quite enough work on hand without such further extensions.

Table 2 also has separate entries for the short and longer terms, and it must be noted that these can conflict. To make the example of the UK, most of us are both emotionally and logically attached to various aspects of 'the British way of life' and want to preserve understatement, cricket, South Kensington, inventiveness, plain speaking, tolerance and many other things that seem to contribute. In the short term, we are threatened by general expectations of consumption that our productive efficiency cannot quite supply. We live in a competitive world in which we are no longer the leaders and pace-setters, though some of

Vol. 7
1979

TABLE 2. SURVIVAL: TOP PRIORITY NEEDS (possible view)

Group	Short term (10 years)	Long term
Human race	Avoid nuclear war	Avoid nuclear war. Maintain supply and exchange
'Western Democracies'	Maintain reasonably free trade	Preserve democracy and freedom
UK	Avoid competitive eclipse	Form necessary alliances to maintain supply

us persist in behaving as though we were. Consequently, we have a short term need to compete successfully according to rules in whose sustainability we may not believe; for this end, we seem to have to behave as though orthodox economic growth can continue indefinitely and energy and raw material supplies can somehow or other go on growing. But if we believe that this 'growth' objective has to be changed or moderated, we are not strong enough to lead a movement to change it now. We then have to win such strength by competing successfully in a game in whose rules we do not believe!

There is no shortage of books on the short term, but they lack long term generality, whether or not they admit this. Our definitive study must insist on the short term survival needs and problems, which are local and sometimes sordid, but must, like Clausewitz, be more especially concerned with greater generality and the longer term.

Technological change has a wide range of different impacts on different groups, and to move beyond the general top-down, synoptic 'survival' imperative, into the bottom-up, synthetic assembly of the top-priority objectives for a particular group. It would be good to take 'the human race' as the group, but this is difficult in a first step because of the great variety and diversity of African, Oriental, and other cultures and thinking. By way of typically British compromise, therefore, I am recommending the group 'Western democracies'.

For this group in the longer term, we have to conduct a subjective selection of priorities to see which impacts of technology we most need to understand. Then we can begin to put together the ideas for the volume "Von der Wende" to stand on the shelf alongside 'Vom Kriege'. I have said that there are many impacts to think about. We therefore have a further need for models, and simple forms of words that summarise situations so as to help with our sorting process. For if every description of a technological impact is complex, convoluted, and expressed in its own jargon, we shall have trouble.

b. *Clarity and simplicity*

Let us take some down-to-earth examples. One technological development about to be started is microelectronic automobile engine management. An automobile or electronic engineer asked to explain this may start by explaining cars, or chips, or combustion. In fact, it will suffice for our purpose to say: 'this is mainly to reduce the amount of unburnt or partly burnt fuel going down the exhaust pipe'. To be sure, there are objections to this very simple statement, but most of them can be answered. We could define the purpose of non-persistent herbicides as being 'mainly to avoid earth movement that is wasteful of energy' or of the synthetic fibres as being 'to produce durable textiles that are easy to look after, and to cheapen and systematise the first stage of textile manufacture by introducing melt 'spinning' or of thermoplastics as being 'mainly to make as many useful objects as possible by low-temperature forming instead of high temperature forming, followed by milling, turning and grinding, thereby saving energy and capital, and reducing cost'. This puts thermoplastic and powder metallurgy technology in contiguous classes, despite the great differences of the methods (injection moulding and hot isostatic pressing) and materials.

One objection to clarity is that clear statement sometimes reveals the weakness of arguments and positions, and no one wants to risk mockery as an unfrocked priest or an unclothed emperor. Moreover, the most important models, generalisations and 'truths' turn out to be so simple that it is difficult to understand how it took all those people all that time to find their way to a statement so very obvious. Philosophers, classicists, and poets have long traditions to back them in the proposition that the delectable distillation of wisdom into a few words can cost lifetimes of blood, sweat and tears. Scientists do not yet have this self-confidence. Sir Peter Medawar may say that simple principle liberates us from the tyranny of the particular so that we need no longer record the fall of every apple, but there are plenty of people whose trade is the extension of the particular, and who do not want to be told that their experiments are superfluous.

Another solid objection to clarity is that this concern with detail is, sometimes (but rarely) entirely justified: was it not the precession of the perihelion of Mercury, in defiance of Newtonian prediction, that played a key part in the development of relativity?

Clear statements very often override or ignore some part of a field, and to be safe about the possible impact of exceptions one has to retreat again into detail and less comprehensible qualification that is the enemy of synoptic thinking and decisive action. In the vast majority of cases, the qualification turns out not to be necessary, but when it is, the consequences of ignoring it can be serious. Being clear, like being decisive, involves a significant risk of important error.

c. *The hierarchy of the effects of technology on society*

Our task—to say what impacts of technology must most importantly be assembled, clarified, codified, and explained in our definitive work—would have been much easier thirty and more years ago. Then, expectation of life at birth was below, and often well below, the scriptural three-score years and ten and life, as well as being short, was often nasty and brutish. Consequently, technology that pushed back famine and disease and thus both prolonged and enriched life, was unequivocally to be welcomed. The production of goods—whether housing (that went with hygiene) or vehicles (that helped develop tractors) or chemicals (that led to pharmaceuticals and agrochemicals) could understandably be accorded social preference over services such as the mending of equipment, cleaning of windows, or collection of rubbish (despite the hygienic aspect). Thus, there was a structured view of 'human progress' which called for technology to be pressed ahead in a particular way. In the West, this view has become obsolete over the last three decades. Now that the expectation of life is over 70 in some countries, and care and comfort is much better, the esteem accorded to longevity and hygiene—and to goods—is being questioned. The extension of physical survival beyond mental survival (so that the decade or so at the end of life is more likely to be spent in insensate senility) is of questionable value. Even when mental death is also postponed, what would be the social effect of survival till (say) 120, so that five generations instead of three have to work out a *modus vivendi*? Thus, longevity and health remain matters where technological impact is of high importance, but it is no longer possible to designate a scheme for technical action that will 'make things steadily better'. Some of the new impact creates unpleasant problems to whose alleviation technology can contribute (e.g. by improving mental function in the old) but where social action is likely to be more important.

In the modern situation, a number of other impacts have emerged that must be given much higher importance than previously. War and conflict is the most obvious case. Technology has always contributed to the destructiveness of war, but social disruption, famine, and disease were usually the major predators. Now, direct destruction is at an unacceptable level for all concerned.

The technological Clausewitz, with his objective 'to survive' can begin by listing the social impacts that he believes simply cannot be neglected over the next four decades. In some cases, remedial technology may continue to be an unalloyed good: resource and energy scarcities may be a case in point, when vigorous technical steps such as the development of solar power can create an important, though not necessarily a sufficient response. In others, technology may seem unlikely to make any important contribution to the solutions; political and social measures being decisively more necessary. In a third class, technology may be simultaneously creating problems and providing means for their solution.

Below this high priority list—which is likely to be embarrassingly long—must come impacts of allegedly lesser effect. Among them will be activities of high appeal to the consumer, but less likely either to solve the key problems or to upset the applecart. And of course, some such cases will acquire new importance unexpectedly. We may therefore expect those concerned with any impact put lower in the list, to urge uprating perhaps because of market appeal, perhaps because of the adage 'just you wait', or perhaps because of ingenuity and spectacle. It is particularly difficult to rate the social impact of the 'gee whiz' operation; the absence of bread may cause riots, but the absence of circuses may help the start of wars. Did the nation that landed men on the moon (with integrated circuitry as a much greater benefit than knowledge of the moon itself) become more peace-loving as a result?

Table 3 is assembled in order to begin the discussion. Its principal object is to distil out

Vol. 7
1979

TABLE 3. SOCIAL IMPACTS OF TECHNOLOGY

1. HIGH IMPACT: PROBLEMS THAT FEED PREDOMINANTLY SOCIAL AND POLITICAL SOLUTIONS

Area affected	Nature of impact	Nature of possible actions
(i) War and Conflict	Scale of destruction and uncertainty of victor's survival	Negotiation and de-escalation
(ii) Democratic process: social bargaining	(a) Longer technological time scales require bipartisan policies	Probably some sacrifice of technical efficiency
(iii) The "economic growth machine"	Satiation of demand where existing products and processes fit need	Public/private sector collaboration in mature industries

2. HIGH IMPACT: CASES WHERE TECHNOLOGY CAN MAKE FURTHER MAJOR CONTRIBUTION

Area affected	Nature of impact	Nature of possible actions
(iv) Resource scarcity (including energy)	Increased bargaining power of resource holders: need for capital: more spoil and waste disposal	More toward conservative technologies: social measures for moderating demand: move to different land areas
(v) Population: health and longevity	Continuing increase (to limit): aggression, senility, new age distributions	Population control: work on mental health of old: social readjustment
(vi) Employment pattern	Move from goods to services: less routine use of brain: some resumed use of muscle	Changed esteem patterns, and some move from specialism
(vii) Food production and preparation	Limited sustainability of Western procedures and diet, for 10 billion people	Conservative husbandry to feed more people per limit area, acceptably
(viii) Environment: toxicology	Negative side effects now large and significant	Proper and co-operative assessment of risk and necessary remedies

3. MEDIUM IMPACT:

Area affected	Nature of impact	Nature of possible actions
Simplification of design	Easier, cheaper production and maintenance: less energy use	Prosecute actively
Substitution and innovation in goods	Longer time scale: higher entry fees: less success	Seek to separate the prosecution of new and established technologies
Size of cities	Telecomms permit reduction (less congestion and aggression)	Involve technology more in civic design
Communications: telephony	Better contacts: less travel: smaller cities	Prosecute actively, especially the psychological and educational components
Information	Cheap processing, retrieval, display	Prosecute actively, on the basis of value analysis
Housing	Improvement	Prosecute actively, with more inclusion of technology and value analysis

4. LOWER IMPACT:

Area affected	Nature of impact	Nature of possible actions
Entertainment	Greater complexity at less cost	Integrate more with education
Speed of transport	More meetings and travel	Integrate with work on telecommunications
Scale of transport and highways	Lower unit capital costs, but more congestion	Integrate with civic design
Navigation	Further improvement	Prosecute actively
Clothing	More role playing	Leave to market mechanism

the high impacts: many more medium and low impacts could be added, so that these groups are no more than illustrative. Each entry in the table calls for comment at much greater length than is possible here. All that can be done is to mention a few highlights, specific and general.

Item (ii) needs more discussion. First, anyone concerned with major technological development knows that continuity of intent, on a

10–20 year scale, is highly desirable. This is too long a period to be acceptable for the normal democratic checks and balances, designed to curb the tendency of power to corrupt. Consequently, there are powerful arguments to take these major developments out of the political arena. Yet, if one then lists all of the matters with long time scales—health policy, medicine, hospital building, education at all levels, design of towns and roads, automobiles and aircraft, new materials and chemicals, control technology, industrial priorities, foreign and defence policy, and so forth, it can be seen that the area remaining for democratic process can be somewhat confined. Secondly, technology has so far tended to create small specialist groups on whom very much larger numbers of others are dependent. Such groups can exercise great economic power: they range from cleaners to systems analysts. It may become necessary to accept some 'inefficiency' in order to reduce the vulnerability of very highly organised and specialised systems—whose 'efficiency' depends on a permanent agreement that no group uses its powerful veto.

Item (iii) also calls for a lot more attention. It is possible to expand the concept of economic growth to include service activities and entertainment, and to use extensibility here to offset reduced employment in industries whose output either does not need to increase (because of stabilisation of demand) or cannot increase (because of rises in raw material or energy costs). It is difficult to predict the dynamics of industrial employment levels, for new industries *do* still arise, though less easily than formerly. What technology certainly creates is an increased need for mid-career retraining and mobility—not easy to achieve in a mature society with high expectation of consumption.

Turning to the areas listed as those of medium and lower impact, virtually all are cases where technology has more to achieve. The case that may generate most argument is that of information. This is undoubtedly an area when all households and individuals can be furnished with quite new levels of access to information and its processing. The question is whether they will want or use this as it becomes available: we already see many times more calculators sold than are used, and books and records bought than are read or played. Regarding industry, the impacts on cost, design and innovation may be considerable. The question is whether the techniques do more than has been considered in the high impact area, under other headings (such as employment).

CONCLUSIONS

Before we summarise the arguments, three matters need attention. The first is the reconciliation of the needs of society for synopsis (generalism) and specialism, and the recognition that here there is a great tension which cannot be eliminated and must be managed. The second is the observation that war itself has proved to be an occasion in which the need for a synoptic view of technological and human capabilities is crucial, so that in a limited way the overall thinking for which we are calling has been done and put into practice. We need to ask whether we would benefit by linking 'Von der Wende' quite explicitly to 'Vom Kriege' so that it can be read as volume three, the connecting volume two being 'Vom technologische Kriege'. The third is the question about what to do on Monday, by way of educating technological generalists who can earn respect and co-operation from the technological specialist.

a. *The synopsis—specialist symbiosis*

Regarding synopsis and specialism, there are several conflicts to be managed or resolved. The first surrounds the risk associated with the unqualified simple statement, already referred to. This conflict cannot be eliminated, and must be accepted and managed. The generalist will be on the side of clarity and acceptance of hazard; the specialist on the side of precision with the acceptance of more or less incomprehensibility. The second is the competition for authority, influence and resources. Both generalist and specialist are necessary, but either can assert that the other is of minor importance, and in a world of limited supplies of the best brains, or research grants, or the time during which the ruler can listen, his claims should come first.

The third concerns method. I recall Michael Polanyi's vivid analogy that "the advancement of knowledge can be likened to the solving of a jigsaw puzzle. The most efficient way to do it is not to organise the players into teams and to issue instructions to them, or to compel

them to work on one particular area of the puzzle: it is to leave each player free to range over the whole puzzle, working individually and uncoordinated with other players".

The specialist is accustomed to the division of labour, and prefers the assembly of the complete picture only after each participant has made his own contribution, with minimum mutual interference and minimum bureaucracy. The problem is that, at the end, there are often problems about the fit of the sections: moreover, it sometimes is found that people have been doing different puzzles. The synoptic generalist probably does the frame first, and gets his friends to help him along the lines of Polanyi's recommendation. But having got his general picture, he needs the specialists if he wants to use it to create and guide practical action. If he is a doctor and decides that his picture indicates a possible brain tumour, he will not be able to take any of the next steps of investigation and surgery himself. If he is a Japanese motor cycle manufacturer and has identified the market need that we mentioned—for a powerful electrical system including a self-starter—he will need new integrated designs of alternator, engine, motor, battery, microelectronics, and so forth—all needing specialist attention as to performance needs, geometry, cost, production technology, and quality control. The specialist often regards the generalist intervention as unnecessarily bureaucratic and tedious, but an assembly that has been 'thrown together' is far less economical and satisfactory than one that has been designed together: usually, of course, the *post hoc* design assembly is impossible.

The fourth concerns freedom to meander. Most really important basic discoveries and inventions have occurred through serendipity, which the specialist therefore reasonably pursues. Yet for each such major advance there must be a multitude of subsequent enabling inventions and design elements, that must be disciplined to make a prototype that works, and extremely disciplined to make a mass production model that is reliable and cheap. Specialists are needed here also. Consequently, the effective generalist has to recognise that some specialists should be very free whilst others must be persuaded, probably against their will initially, to combine and accept synthesis that again will be seen as bureaucratic.

With all of these problems, there can be no surprise that the vitally necessary reconciliation of synopsis and specialism is difficult, or that specialist science in Britain is resistant to synoptic concepts. One obstructive device is simply to say that anyone who is promoting synopsis is automatically attacking specialism. This is emphatically not so for me: in urging attention for synopsis, for technology, and for engineering design, I am in no way attacking specialism of any kind.

b. *The synopsis of technology and sociology in war*

We now turn to the significance of war as an environment where there must be some kind of technological synopsis that effectively integrates all possible specialisms. Some interesting cases are very broad subjects, such as the feeding of an entire population under blockade, which may require the integration of rationing, studies of nutrition, programmes of shipbuilding, and the technology of submarine detection and destruction; everything affects everything else.

Others are sharp and narrow such as the building of an atomic bomb, whose impact would be so spectacular that there is little initial need to give much consideration to delivery systems: one will be found. But the important overall finding is that this mechanism can sometimes develop technology in a way that peacetime procedure has failed to do. A case in point is food production: both of the two world wars caused the performance of British agriculture to turn upward discontinuously—not because of new invention, but because of the simultaneous application of many possibilities that had been made available, but not used, in peacetime. Another is the gas turbine, that had often been discussed in peacetime, but never built. Nuclear energy release and polyethylene were both invented in peacetime and developed in wartime. Then, after the war, the new technology is available at a cost that can be built into the peacetime economy.

Now that entry fees are more often too high for the traditional institutions, it is vital to find methods for efficiently assembling enough money (probably tax revenue) for the purpose, and to roast the first pig without burning down the house. The Apollo programme pulled through a great deal of integrated circuit, com-

puter systems and software technology. How can we, for example, pull through the many skills needed to make a lightweight, non-corrodible, energy-economic car bodyshell made from fibre-reinforced thermoset? More generally, how can we cheapen and control the properties of plastics so as to enable them to replace metals in engineering uses (as opposed to simpler and less critical duties in packaging, roofing, and domestic hardware)? This is a task to begin on Monday, not in the year 2000.

c. *Summary*

1. In Britain until recently, the politician, the businessman, the engineer, and the scientist have been able to avoid detailed mutual interaction, interference, and understanding. General devices, such as the patent and limited liability systems have sufficed.

2. The increasing scale, penetration, and financial and environmental impacts of technology have raised problems insoluble by the traditional general procedures. More detailed and specific interactions seem to be needed.

3. The initial steps have been useful but defensive; people have reluctantly begun to learn about the skills and tasks of other groups, mainly to preserve their own effectiveness, and if possible their traditional freedoms. Engineers have learnt some economics, and politicians and businessmen have learnt more about technology.

4. The problems of interpenetration, however, continue to increase, so that synoptic policies (such as the national industrial strategies) seem to be called for. This requires synoptic understanding that threatens specialist autonomy.

5. Countries other than UK have faced these needs earlier, some (e.g. USA) over a century ago.

6. There is a superficial and somewhat bizarre analogy between the impacts of technological change and war, even though one is a positive-sum and the other a negative-sum game. The synoptic practise of war was invented by Alexander, Napoleon and others, and codified clearly by Clausewitz.

7. Synoptic practice of technological change has now begun, and clear codification, following Clausewitz, is needed.

8. Now we require a great increase in clarity, simplicity of perception and demystification which can begin at once. It will, however, be unpopular because it will reveal unclothed emperors.

9. We also need a view on the sensitivities of social change to the many impacts of technology, and on priorities.

10. These priorities will need much discussion, but impacts on democracy, the role of the individual, the veto of small specialist groups, the rapid growth of resource demand in relation to often diminishing resource supply, and changes in environment, congestion and aggression seem crucial.

11. A flight from technology cannot solve the problems without a catastrophic drop in population.

12. Synoptic generalism is symbiotic with specialism: each needs the other, but they compete for resources and involve philosophical tensions that must be managed.

ADDRESS FOR CORRESPONDENCE: *Duncan Davies, Esq, Chief Scientist and Engineer, Department of Industry, Abell House, John Islip Street, London SW1P 4LN, UK.*

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 4, pp. 287 to 295
© Pergamon Press Ltd 1979. Printed in Great Britain

0306-0483/79/0801-0287$02.00/0

Liquidity Planning in a Small Bank

DALEN T CHIANG

Cleveland State University, Ohio

(*Received August* 1978; *in revised form January* 1979)

Given a forecast of supply and demand for cash in each period of an infinite planning horizon, and with a known current portfolio, a policy is chosen to invest these cash supplies in securities of different maturities so that the demand in every future period can be satisfied by securities maturing in that period. The objective is to maximize the minimum of the excess over the planning horizon so that any illiquidity in one period is spread out over the entire planning horizon. Analytical solutions are obtained for single maturity and barbell investment policies. Feasibility and optimality conditions are determined for these policies.

1. INTRODUCTION

A BANK is an institution that plays a role as financial intermediary in the economy, its two major activities being borrowing and lending. Typically, a bank accepts money from depositors, lends a portion of these deposits to borrowers, and, by the art of brokerage, may profit from a spread between lending and borrowing rates. To support such operations, banks also maintain other activities; among these liquidity planning is an essential one. As Komar [12] points out, "the primary purpose of commercial bank liquidity planning is that of supporting other banking functions, by providing insurance that deposit run-offs are readily met, and by maintaining inventory for lending activities".

Liquidity planning in a small bank completely differs from liquidity planning in a large bank basically because of the ability that a large bank can purchase funds, such as certificates of deposit (CD), Eurodollar and federal funds, whenever needed. Smaller banks cannot rely their liquidity needs in the liability management for when they need money the most it is usually at a time when CD market is the tightest and the interest rate is the highest. In fact, most small banks depend solely on the management of assets to provide their liquidity needs.

Briefly, we will discuss some of the general practices of liquidity planning in small banks. In our discussion, we will restrict our attention to those aspects directly related to a bank's liquidity planning, which although is only a part of bank's overall investment planning. Other factors such as pledging requirement, risk and tax considerations which are important in bank's investment planning but are not essential in liquidity consideration will not be discussed in this paper.

The first step in liquidity planning is to estimate liquidity needs. A small bank's need for liquidity essentially consists of two components; to meet deposit withdrawals and to satisfy the legitimate credit need of the community. Thus, deposit level and loan demands must be forecasted. For most small banks, deposit level tends to follow a seasonal pattern with certain long term trend. Hence forecast on deposit level is relatively easy. Depending on the economy of a bank's local community, loan demands are much harder to forecast, because, unlike deposit level, loan demands contain much more cyclical variations. However, banks do have more control over their loan portfolio than deposit accounts, and in a time of tight credit, loans are granted only to prime customers.

The second step is to determine the size of the investment portfolio. Nearly all small banks used to accept the size of their investment portfolio as a residual. However, the tremendous growth in loans that many small

banks have experienced in the recent years and the fact that liability management is not a practical source of liquidity gradually change the residual nature of the investment portfolio. The size of the portfolio is determined according to the estimation of bank's liquidity needs.

Once the size of the portfolio is determined, a bank has to set a maturity policy, which involves the establishment of a maximum maturity limit and the scheduling of maturities within the portfolio. The composition or distribution of maturities in a portfolio is referred to as its *maturity structure*. A portfolio has *spaced maturity structure* if the maturities of securities are more or less spaced evenly within the established maximum range. The proceeds of maturing securities are reinvested in securities of the longest maturity desired in the portfolio. In this way, a bank which has a bond portfolio with spaced maturity structure is not gambling on changes of interest rates or state of the economy. The yield of the portfolio is at the rate of the longest maturity while the average maturity is only half of the longest maturity. As long as the yield curve is increasing, the spaced maturity structure will assure the bank of higher-than-average income. In addition no active trading is needed to maintain a spaced maturity structure. Many small banks prefer the spaced maturity structure in their bond portfolio [3, 5, 11, 15].

Finally, a bank has to determine strategies on the purchasing, trading and switching of bonds for the investment portfolio. There are important financial decisions to be made, such as risk, return and diversification, which are not discussed here. Most small banks buy securities for their investment portfolio and put them away until they mature. The main reason is that officers in a small bank usually do not have the time and/or expertise to actively trade investment portfolio. With a maturity strategy such as spaced maturity structure, small banks do not have to actively trade their investment portfolio.

In this paper, we are interested in the problem of selection of portfolio's maturity structure. This problem and some related problems have been studied by Bradley and Crane [1–3, 7], Cheng [4], Chiang [5, 6], Watson [15], Wolf [16] and Woodsworth [17]. In particular, Bradley and Crane are interested in the performance of a portfolio with spaced and barbell maturity structures. A portfolio has a *barbell maturity structure* if a portion of the portfolio is in short-term securities and the remainder is in long-term securities. The advantage of a barbell maturity structure is that the short-term portion provides liquidity while the long-term portion provides returns, based on the assumption that the yield curve is usually increasing. Obviously, frequent tradings must be made in order to maintain a barbell maturity structure.

An alternative approach to maturity structure of a bond portfolio is proposed in this paper. Instead of considering the maturity structure of an entire portfolio, we study the maturity composition of each investment decision. For each investment, the *investment policy* specifies the fraction to be investigated in each maturity. The motivation of studying investment policy is that since many small banks do not actively trade their bond portfolio and since these are in general kept until maturity, it follows that the change of a portfolio's maturity structure depends only on the distribution of maturities of the newly invested portion of the portfolio. Hence, it is rather natural to consider the problem of scheduling maturities of each investment. Furthermore, when an investment policy is given, the maturity structure of the entire portfolio could be easily computed by an aging equation, which will be discussed in the next section.

More specifically, in this paper we are interested in the following asset liquidity planning problem: In each time period, a certain known cash supply is available for investment in government securities of various maturities. Forecasts of cash demand are assumed to be given for each planning period. We seek investment policies such that all cash demands in later periods can be met by matured securities, and in addition, the risk of illiquidity caused by unforeseen cash demands is minimized over the entire planning horizon. Thus, our objective is not to maximize the yield, instead, we want to share liquidity among all planning periods.

In the next section, we formulate the asset liquidity planning model. The model is to maximize the minimum excess over an infinite planning horizon subject to a profit constraint, where the *excess* is defined to be the dollar amount of matured securities less cash

Vol. 7
1979

demands in each period. Then the model is used to analyze investment policies under the steady state condition in Section 3. In particular, we are interested in single maturity and barbell investment policies. A *single maturity investment policy* invests the entire amount in securities of the same maturity. An investment policy is *barbell* if a portion is invested in very short term securities and the remaining in securities of a longer maturity.

We show that a single maturity policy is optimal if matured securities from the present portfolio is larger than certain amount in each period before the initial investment is matured. It is shown in [6] that when investment policy is single maturity, the maturity structure of the portfolio will be spaced. Furthermore, if the present portfolio has a spaced maturity structure, single maturity policy will maintain the spaced maturity structure. Thus, the model has explained the intuition of current practice of many small banks that, barred from any sudden demand for a large quantity of cash in the near future, maintaining a portfolio with spaced maturity structure is optimal in the context of sharing liquidity.

When the optimality condition for single maturity investment policy is not satisfied, then either the profit constraint is violated or there is a time period with relatively low excess. Depending on the situation, two types of barbell policies should be considered. One type of barbell policy improves liquidity, while the other improves yield. We show that under appropriate conditions, these barbell policies are optimal. In the last section, we will give some additional remarks on the implementation of our results and directions for future research.

2. DEFINITIONS, ASSUMPTIONS AND THE MODEL

Let $t = 1, 2, 3, \ldots$, denote time periods. For each period t, a *supply* s_t is allocated for investment. A *demand* for cash d_t is forecasted. Both s_t and d_t are nonnegative. Without the loss of generality, we assume that m-period is the maximum maturity allowed in the bond portfolio. It follows from this assumption that there are at most $m + 1$ different 'types' of securities in the portfolio. A security is of type 0 if it matures within the period of purchase. A security which is purchased at the beginning of a period t is called type j, $j = 1, 2, \ldots, m$, if it matures at the beginning of the period $t + j$. In other words, a security is of type j if it matures j periods after its purchase date.

We also assume that there is no sale before maturity. To illustrate, if a security of type 4 is purchased at the beginning of the first period, and is expected to be sold at the beginning of the fourth period, then this particular security is of type 3 instead of type 4, because it 'matures' three periods following the date of its purchase. We state that on the initial date of investment, no sale before maturity is expected.

Thus, type j securities represent those securities which have a remaining life of j periods, while type 0 securities in a period simply represent those securities which will mature within that particular period. A type j security in period t will become type $j - 1$ security in period $t + 1$.

A *stationary investment policy* $X = (x_j: 0 \leq j \leq m)$ allocates a fraction x_j of supply s_t to be invested in securities of type j at the beginning of each period t for each j. Thus, a single maturity investment policy is $x_k = 1$ for some k with $0 \leq k \leq m$, and $x_j = 0$ for all $j \neq k$, and a barbell investment policy is $x_0 > 0$, $x_4 > 0$, for some k, and $x_j = 0$ of $j \neq 0$, k. The motivation for barbell investment policy is that the portion invested in the short end provides liquidity, while the remaining portion seeks yield. The advantage of stationary policies is their simplicity, although the assumption may be somewhat unrealistic. The mechanical trading rules in Bradley and Crane [3] are stationary policies. Because of the lack of resources to obtain investment expertise, many small banks employ investment policies which are almost stationary. In this study, we will only consider stationary policies.

For a stationary investment policy, we assume that there is no short sales since a short sale is a liability rather than an asset. This assumption is equivalent to the following nonnegativity constraints:

$$x_j \geq 0 \qquad 0 \leq j \leq m. \tag{1}$$

Usually, only government bonds are included in the portfolio, and thus (1) also excludes any bond which can be issued by a bank. Since cash can be considered as a type 0 security,

there is no loss of generality by assuming s_t is fully invested for each period t, i.e.

$$\sum_{j=0}^{m} x_j = 1 \tag{2}$$

A portfolio in period t, $t = 1, 2, \ldots$, is represented by $Y_t = (y_{j,t}: 0 \leqslant j \leqslant m)$, where $y_{j,t}$ is the dollar amount of type j securities in period t. Recall that the index j refers to the remaining lifetime in periods of securities in a portfolio. The period before the first planning period is called period 0, which is usually the current time period. The portfolio in period 0 is called the *current portfolio*, and is denoted by Y_0.

Aging equations

In order to have a conceptual understanding of the age and maturity structure of an asset portfolio, we discuss how assets age with the passage of time, and how a collection of such assets leads to a maturity structure for the portfolio as a whole. For this purpose, we derive a mathematical relationship between a period-t portfolio and a stationary investment policy.

A portfolio in period t is the collection of securities inherited from period $t-1$ and new securities purchased in period t. Since a security of type $j+1$ in period $t-1$ becomes a type j security in period t, the dollar amount of type j securities inherited from period $t-1$ is $y_{j+1,t-1}$ for $0 \leqslant j \leqslant m-1$. The dollar amount of type j securities purchased in period t is $s_t x_j$. Hence, it follows that the dollar value of securities of type j in period t is

$$y_{j,t} = y_{j+1,t-1} + s_t x_j \qquad 0 \leqslant j \leqslant m-1$$
$$y_{m,t} = s_t x_m,$$

which is the one period aging equation.

Given a current portfolio and a sequence of cash supplies, s_t, $t = 1, 2, \ldots$ we are interested in the maturity structure of the period-t portfolio when an investment policy (x_j) is employed. By repeated recursions of the one period aging equation we get

$$y_{j,t} = y_{j+t,0} + \sum_{i=1}^{t} s_i x_{j+t-i} \qquad 0 \leq j \leq m, \tag{3}$$

where $y_{k,0} = 0$ and $x_k = 0$ for all $k > m$. The aging equation (3) shows that type j securities in period t are the sum of type $j+t$ securities in the initial portfolio and a fraction of all securities invested in later periods.

Profit constraint

In the model we include a profit constraint to ensure that all feasible investment policies achieve a certain minimum level of profitability. In our context, the word profit is defined to be the discounted present worth of all future yields from investments, including coupons and capital gains resulting from buying at discount and selling at par.

Because of the time value of money, profits from two investments with different payment schedules may be different even when the interest rates are identical. There are usually three different types of payment schedules: purchasing at discount and selling at par, payment of interest at maturity and periodic coupon payments. The first two are identical mathematically: purchasing a security of par value u dollars at the discount price of v dollars is equivalent to investing v dollars in a security and receiving a total payment of $(u - v)$ dollars together with the principal v dollars at maturity. Thus, we will assume that all securities have either periodic coupon payments or payment at maturity. In the formulation of our profit constraint, one must include securities with both types of payment schedules as most long-term bonds have periodic coupon payments, while most short-term securities are paid at maturity. In the Appendix, we show that in either case, we can define a quantity r_j such that r_j represents the present value of total future return from investing s_t in type j securities at period t. The present value of the total future return of a policy is

$$\sum_{j=0}^{m} r_j x_j,$$

which we will require to be larger than a predetermined profit level p. In other words, the *profit constraint* is

$$\sum_{j=0}^{m} r_j x_j \geqslant p. \tag{4}$$

A policy is *feasible* with respect to the profit constraint if it satisfies the inequality (4). If $r = \max(r_j)$ and if $p > r$, then no policy is feasible with respect to the profit constraint. The profit constraint is *redundant* if every policy is feasible with respect to the profit constraint.

If $p < r$, then p divides securities into 'profitable' and 'unprofitable' types,. For a policy to

Vol. 7
1979

be feasible with respect to the profit constraint, the fraction to be invested in 'unprofitable' securities is bounded from above, and at least one 'profitable' security must be purchased.

Objective function

If the dollar value of securities maturing in one period is greater than the cash demand in that period, then the demand can be easily satisfied. If the value of securities maturing in period t is $y_{0,t}$ and the demand is d_t, there is an excess e_t when $e_t = y_{0,t} - d_t > 0$. From (3),

$$y_{0,t} = y_{t,0} + \sum_{i=1}^{t} s_i x_{t-i}.$$

Our goal will be to satisfy all cash demands with funds derived from matured securities. Obviously, all demands will be satisfied if the excess is nonnegative for all periods. This occurs if and only if the minimum excess (over all periods) is nonnegative. The objective we will use is to find an investment policy such that the minimum excess is maximized, i.e.

$$\max_X \left(\min_t (e_t) \right). \tag{5}$$

The maximization in (5) is over all policies X which satisfy the profit constraint (4), the full investment constraint (2) and nonnegativity constraints (1). Since

$$e_t \geq D \quad \text{for all} \quad t, \tag{6}$$

the max–min problem of (5) can be solved by the following linear program (Dantzig [8], Chapter 13):

$$\max D \quad \text{subject to (1), (2), (4) and (6).} \tag{7}$$

where $y_{t,0} = 0$ for all $t > m$. The inequality (6) is called the *cash flow constraint of period t*.

The quantity D has several interpretations. First, it can be regarded as the minimum excess over the entire planning horizon. This interpretation arises from the fact that the left hand side of (6) is the excess in period t; we require this excess to be greater than or equal to D in every period. If a feasible solution of the linear program (7) with $D \geq 0$ exists, then D must be nonnegative in every optimal solution of (7). Thus, every period has a nonnegative excess, and sales of securities before maturity, (i.e. at discount) will not be necessary.

We may interpret $-D$ as borrowings, which is more meaningful when D is negative, that is, when a deficit occurs. Since maximizing D is equivalent to minimizing $-D$, the objective function of (7) can also represent the minimization of borrowings. D can also represent a safety stock for unexpected demands. The maximization of D can then be interpreted as finding the largest safety stock to meet unexpected demands.

3. OPTIMALITY IN EQUILIBRIUM

In this section, we present necessary and sufficient conditions for feasibility and optimality of single maturity and barbell investment policies when demands and supplies are stationary. Complete mathematical proofs of feasibility and optimality conditions will not be included in this paper (for rigorous mathematical proof, see [5], chapter 3). However, the intuition behind these conditions will be discussed.

The stationary model

In equilibrium, supply and demand are assumed stationary, i.e. $s_t = s$ and $d_t = d$ are constant for all t. With such assumption, the cash flow constraint (6) becomes

$$e \geq D \tag{8}$$

The linear program (7) becomes

$$\max w \tag{9}$$

subject to (1), (2), (4) and (8).

Each investment policy corresponds to one or more basic solutions of the linear program (9). Since all non-basic variables are zero, a basic solution is found by solving the constraints as a system of simultaneous equations. Feasibility conditions insure the nonnegativity of all variables, including slack variables, and conversely. Hence, the feasibility condition of a particular basic solution can be derived from the nonnegativity of all basic variables in that solution. Furthermore, in a basic feasible solution, if $w = s$, then the solution is already optimal because of the constraint $w \leq s$. If $w < s$, one can determine the optimality conditions from the duality theory of linear programming (Dantzig [8], Chapter 6), in particular, from the complementary slackness condition and the strong duality theorem.

Single maturity investment policy

Consider the following single maturity

investment policy:

$$x_k = 1 \qquad \text{for some} \quad k \leqslant m \tag{10}$$
$$x_j = 0 \qquad \text{for all} \quad j \neq k.$$

For this single maturity policy to satisfy the profit constraint, the return from type k securities must be larger than the predetermined profit level p.

For the policy (10) to satisfy cash flow constraints (8), we note that a basic feature of the single maturity investment policy is that the investment in the first period will not mature until k periods later. Hence, for the first k periods, demands must be satisfied by cash generated from matured securities in the current portfolio, i.e. $y_{j,0}$ where $j \leqslant k$. After period k, the investment from the first period will mature, hence for each period $t > k$, the total amount matured will be at least s. Yet $w \leqslant s$ is one of the constraints, thus, if $y_{j,0} \geqslant s$ for all $j \leqslant k$, w could equal to s and thus achieving optimality, providing that the profit constraints is also satisfied, i.e. $r_k \geqslant p$. In this case, the minimum excess D is equal to $s - d$.

We should point out that whenever there are more than one type of security which satisfies the optimality condition, the one with the shortest maturity would be preferred. This is to say that liquidity is preferred in terms of shorter maturity when the profit level is satisfied. This corollary essentially follows from the objective function of the model.

Barbell investment policy

The advantage of a barbell policy is that the short end provides liquidity while the long end provides yield. Thus, we can assume without the loss of generality that

$$r_0 \leqslant r_m = \max_{0 \leqslant j \leqslant m} r_j,$$

for we should not consider securities with longer maturity and yet provide less profit. With this assumption, the profit constraint would be redundant for all barbell policies if the profit level $p \leqslant r_0$. If $p \geqslant r_m$, then no feasible barbell policy exists. Therefore, to study the optimality of barbell policies, we only consider the case of $r_0 < p < r_m$. Optimal mix between the fractions of short (x_0) and long (x_m) maturities must be found, so that the maximum liquidity is obtained while the profit level p is achieved. This balance between profitability and liquidity will depend on the profit level p, the future returns r_0 and r_m, and the current portfolio Y_0.

Since $r_0 < p$, type 0 security is an unprofitable type, thus x_0 is bounded above by $\beta_0 = (r_m - p)/(r_m - r_0)$. This upper bound suggests the consideration of the following barbell policy:

$$x_0 = \frac{r_m - p}{r_m - r_0}$$
$$x_m = \frac{p - r_0}{r_m - r_0} \tag{11}$$
$$x_j = 0 \quad \text{for all} \quad j \neq 0, m$$

The reason is that among all barbell policies which are feasible with respect to the profit constraint, policy (11) allocates the largest fraction of each period's supply to type 0 securities, and thus, is the most liquid one.

The cash flow constraints (8) under any barbell policy can be rewritten as

$$\begin{aligned} w - sx_0 &\leqslant y_{j,0} \qquad \text{for} \quad 1 \leqslant j \leqslant m \\ w - sx_0 &\leqslant sx_m \end{aligned}$$

since $x_0 + x_m = 1$. Thus, a barbell policy can be considered as the single maturity policy of investing a supply of sx_m in type m securities given that a demand is $d - sx_0$. From the optimality theorem of a single maturity investment policy, a feasible barbell policy is optimal if $y_{j,0} \geqslant sx_m$ for all $j \leqslant m$, i.e.

$$x_m \leqslant y_{i,0}/s, \tag{12}$$

where $y_{i,0}$ is defined by

$$y_{i,0} = \min_{1 \leqslant j \leqslant m} y_{j,0}. \tag{13}$$

Hence, a barbell policy is optimal if it satisfies both (12) and the profit constraint. In particular, the barbell policy (11) is optimal provided that $(p - r_0)/(r_m - r_0) \leqslant y_{i,0}/s$.

When $(p - r_0)/(r_m - r_0) > y_{i,0}/s$, it can be shown from the optimality theorem of a single investment policy and the duality theory of linear programming that the barbell policy (11) is still optimal if the condition $r_0 \geqslant r_j$ for every $j < i$ is satisfied. This condition, which means that type 0 securities have higher yield than type j securities for all $j < i$, is probably unlikely in practice. Nevertheless, when this situation happens, investments in type 0 security

Vol. 7
1979

are not only for its short maturity but also for its yield.

There is an alternative optimal barbell policy when the condition $(p - r_0)/(r_m - r_0) \leqslant y_{i,0}/s < 1$ is satisfied. The barbell policy

$$x_0 = 1 - \frac{y_{i,0}}{s}$$

$$x_m = \frac{y_{i,0}}{s} \tag{14}$$

$$x_j = 0 \quad \text{for all} \quad j \neq 0, m,$$

is optimal because (i) the inequality $(p - r_0)/(r_m - r_0) \leqslant y_{i,0}/s$ insures that the profit constraint is satisfied, (ii) $y_{i,0}/s < 1$ guarantees the nonnegativity of variables, and (iii) for the basic solution (14) the objective value of linear program (9) is s.

As we have pointed out earlier, a barbell policy provides an excellent example of the tradeoff between profitability and liquidity. For the policy (11), the fraction invested in long-term securities would be increased to satisfy a higher profit level, while the portion of short-term investments would have to be reduced. When p decreases, the requirement for profits is reduced and short-term liquid investments increases. The policy (14) is more concerned with the amount of securities maturing in period i. If the amount of securities maturing in period i decreases, then the portion invested in short-term securities must be increased to provide liquidity in period i.

4. FURTHER IMPLICATIONS

The goal of our model is to provide a theoretical framework from which long range investment policies of liquidity planning in small banks can be analyzed. The emphasis is on the analysis of long range investment policies, i.e. stationary policies. In particular, we are interested in sufficient conditions such that these policies are optimal in the sense that resources are shared equally among all planning periods to satisfy demands. For a bank with sharing liquidity as its objective in liquidity planning, an investment policy could be implemented if the sufficient conditions for that policy are satisfied.

The equilibrium condition is very restrictive as far as a bank's short term liquidity management is concerned. However, for many small banks, the long term trends in both deposit level and loan demands (as measured by the size of their loan portfolio) are linear. If we consider a year to be one period, then supply and demand would be very closed to the steady state condition, and hence, the results from the last section can be used directly. The advantage here is that no algorithm is needed.

For short term liquidity management, the presence of seasonality in deposit level and loan demands makes the steady state assumption unrealistic, and thus the equilibrium results are not applicable. In this case, one could solve the infinite linear programming model (7) in Section 2 by algorithms similar to the one proposed by Grinold [9]. Alternatively, one could formulate finite planning horizon models as that of Komar [12]. However, analytical solutions could not be found for any nonequilibrium problem. In addition, since the planning horizon is finite, careful considerations must be given to the ending portfolio so that the liquidity provided by the ending portfolio for those time periods beyond the planning horizon would be sound.

In our planning model, the max–min objective of sharing liquidity only concerns with cash flows in each planning period. Since interest rates are not included in the objective function, investment policies are usually insensitive to the change of interest rates. Previous models on the management of government bond portfolio have profit maximizing objective and thus are very sensitive to the forecasts of interest rates (see Watson [15]). While accurate prediction of future interest rates is hard, it is much easier to forecast supply and demand of future cash flows for many small banks. Thus, insensitivity to future interest rates is an important advantage of the model presented in this paper.

To forecast exact future cash flows may often be a complex task. However, it is not necessary to have accurate forecasts of exact future cash flows in order to implement the results of the last section. For example, we could consider a period to be one year in the long range planning and thus future supply and demand are quite aggregate. With stable long-term trends and predictable short-term seasonality in both deposit level and loan demands in many small banks, forecasts for future cash flows in annual

or semi-annual intervals are much easier in these banks. In addition, equilibrium solutions in the last section are not very sensitive to the data on supply and demand. For example, if we assume that the profit constraint is not active, the optimality of single maturity policies only depends on the size of each type of securities in the current portfolio as measured by the equilibrium supply. Thus, if type j securities in the current portfolio are much larger than the supply for all $j \leqslant k$, then the single maturity policy $x_k = 1$ will be optimal as long as the supply is not under-estimated by a big margin. Furthermore, the size of the supply does not solely depend on forecasts of future cash flows, but also partially depends on the management's allocation of resources, which is a part of a bank's overall asset liability planning and is not included in our liquidity planning model.

From the above discussion, several directions of future research emerge. First, investment policies different from single maturity and barbell policies should be investigated and evaluated. Second, the stochastic nature of supply and demand could be included in liquidity planning models. Third, one should formulate liquidity management models (vs liquidity planning models in this paper) which could incorporate tradings of securities and capital gains and losses resulted from these tradings and interest rate fluctuations.

A different direction is to apply the approach presented in this paper to liability liquidity. To be sure, asset liquidity is and should be considered as only one component of a bank's overall liquidity planning. A bank should have an overall liquidity planning program, rather than relying solely on either asset or liability management for its liquidity needs. This is pointed out also by Ritter [14] after his study of 1966 credit squeeze. However, we can model asset and liability liquidity separately, and then incorporate them into an overall liquidity planning program. This paper is intended to be the first step toward an overall liquidity planning model.

ACKNOWLEDGEMENTS

This research has been partially supported by a fellowship from Bank Administration Institute. The author wishes to thank Robert M Oliver of University of California at Berkeley, Jack A Brollier, Bertwing C Mah and John R Paulus of the California State Banking Department for many valuable discussions. The author would also like to thank the Chief Editor and a referee of this Journal for helpful suggestions.

REFERENCES

1. Bradley SP & Crane DB (1972) A dynamic model for bond portfolio management. *Mgmt Sci.* **19** (2), 139–151.
2. Bradley SP & Crane DB (1973) Management of commercial bank government security portfolio: an optimization approach under uncertainty. *Jl Bank. Res.* **4** (1), 18–30.
3. Bradley SP & Crane DB (1975) Simulation of bond portfolio strategies: laddered versus barbell maturity structure. *Jl Bank. Res.* **6** (2), 122–134.
4. Cheng PL (1962) Optimum bond portfolio selection. *Mgmt Sci* **8** (4), 490–499.
5. Chiang DT (1976) *A longitudinal planning model of asset liquidity for commercial banks.* ORC Report 76-14, Operations Research Center, University of California, Berkeley, USA.
6. Chiang DT (1978) *Investment policies and maturity structure of a bond portfolio.* Working Paper 78-1. James J Nance College of Business Administration, Cleveland State University, Cleveland, Ohio, USA.
7. Crane DB (1971) A stochastic programming model for commercial bank bond portfolio management. *Jl Fin. Quant. Analysis* **6** (3), 955–976.
8. Dantzig GB (1963) *Linear Programming and Extensions.* Princeton University Press, Princeton, New Jersey, USA.
9. Grinold RC (1977) Finite horizon approximations of infinite horizon linear programs. *Math. Programming* **12** (1), 1–17.
10. Grinold RC, Marshall KT & Oliver RM (1976) Longitudinal manpower planning models. *Nav. Res. Logist. Q.* **23** (2), 245–260.
11. Hempel GH (1975) 14 steps to a sound investment policy for a smaller bank. *Banking* **67** (2), 38–43.
12. Komar RI (1971) Developing a liquidity management model. *Jl Bank Res.* **2** (1), 38–53.
13. Pulliam KP (1977) A liquidity portfolio management strategy approach. *Jl Bank. Res.* **8** (1), 50–58.
14. Ritter LS (1967) *Bank Liquidity Re-Examined.* Association of Reserve City Bankers.
15. Watson R (1973) Tests of maturity structures of commercial bank government securities portfolio: a simulation approach. *Jl Bank. Res.* **3** (1), 34–46.
16. Wolf CR (1969) A model for selecting commercial bank government securities portfolios. *Rev. Econ. Stat.* **51** (1), 40–52.
17. Woodworth GW (1967) *The Management of Cyclical Liquidity of Commercial Banks.* The Bankers Publishing Company, Boston, Massachusetts, USA.

Appendix

We compute the present value of total future return from type j securities which either have periodic coupon payments or payment at maturity. Let the discount factor per period be δ, where $0 < \delta < 1$. Suppose one dollar invested in a type j security in period t results in a return of $r_{j,t}$ at maturity. Then the present value of returns from the investment of a fraction x_j of s_t in the security is $\delta^{j+t-1} r_{j,t} x_j s_t$. Summing over all future periods the present value of future return (profit) from type j

securities is

$$\delta^j\left(\sum_{t=1}^{\infty}\delta^{t-1}r_{j,t}s_t\right)x_j.$$

If coupon payments of a type j security purchased at period t are made in each period, then coupons are received in periods $t, t+1, \ldots, t+j-1$. If the coupon rate is $\bar{r}_{j,t}$ per period paid at the end of each period, it follows that the present value of these coupon payments from an investment of $x_j s_t$ in the security is $((1-\delta^j)/(1-\delta))\delta^t\bar{r}_{j,t}x_j s_t$. Hence, the present value of total future returns from type j securities is

$$((1-\delta^j)/(1-\delta))\left(\sum_{t=1}^{\infty}\delta^t\bar{r}_{j,t}s_t\right)x_j.$$

Thus we define

$$r_j = \delta^j\left(\sum_{t=1}^{\infty}\delta^{t-1}r_{j,t}s_t\right)$$

if the payment of type j securities is received at maturity, and

$$r_j = ((1-\delta^j)/(1-\delta))\left(\sum_{t=1}^{\infty}\delta^t\bar{r}_{j,t}s_t\right)$$

if periodic coupon payments are received for each dollar invested in type j securities. That is, r_j represents the present value of total future return from investing s_t in type j securities at period t.

ADDRESS FOR CORRESPONDENCE: *Dalen T Chiang, Esq, College of Business Administration, Cleveland State University, Cleveland, Ohio* 44115, *USA.*

ol. 7
979

BLANK
PAGE

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 4, pp. 297 to 308
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0801-0297$02.00/0

An Integrating Model of Motivated Work Behaviour

MARTIN WALBANK

Institute of Science & Technology, University of Manchester

(*Received April* 1978: *in revised form March* 1979)

A model of the processes of work motivation is presented which integrates concepts of trait, reinforcement and valency-expectancy theories. The response is determined by a three level structure; a continuing evaluation process, a set of learned constructs and a set of highly stable personality dimensions. The motivated response, in working situations, is considered as a series of discrete behaviour choices; the dynamics of the behavioural response can be established from the information processing and learning processes within a feedback connected model.

THE STUDY of people in working situations and of the relationships between task characteristics and workers' responses to social and organisational pressures forms one of the main areas of current research. The belief that work can and should be structured to provide satisfaction as well as performance has become a major theme in both experiment and management thinking over the last two decades.

In much of the recent work there may be an element of over-reaction to the earlier mechanistic 'Tayloristic' approaches to the design and management of jobs [28].

Judged by criteria of replication, diffusion of successful experiments within organisations, widespread application and permanence of results, research to date generally falls well short of the standard required for the design of practical motivational and work design programs. Current evidence relating motivation to performance and satisfaction in work situations is widespread yet often inconclusive or contradictory [9, 21, 33, 36]. Despite the level of interest and amount of research, there is surprisingly little hard evidence for convincing working groups and managements of the existence of sound theoretical models of the determinants of work behaviour which can be used to re-structure work leading to predictable and stable improvements in desired performance, satisfaction and quality of working life.

Work motivation concerns the process by which choices are made by individual subjects, as they satisfy currently perceived needs, from the forms of activity which are made available within the social, technical and organisational environment of the work situation.

"A theory of motivation should be able to determine why some outcomes are sought and the factors that influence how they are sought" Lawler [20].

'Blanket' theories ignoring the subjects' individual differences and the interactions between them and their environments are likely to be of little operational utility.

Sound theoretical models defining the factors and interactions which determine workers' responses to the structure of working situations are a necessary foundation for the design of experimental studies of human performance and, for the development of effective prescriptive change programs seeking to improve both performance and satisfaction by the redesigning of jobs.

TIME RESPONSE MOTIVATION DYNAMICS

For most people work is a continuing activity and an understanding of the patterns of motivational response over time must form a key element of any satisfactory motivational

theory. The dynamics of learning and extinction were extensively studied during the development of drive and reinforcement theories [19, 37]. Although Lawler [20] and Dachler & Mobley [10] have proposed models including feedback loops and learning there appears to have been little interest in the dynamics of motivation and experimental evidence relating to the variations of motivational responses over time is rare. If the motivational process in working situations involves learning and feedback processes, cross sectional studies will provide only limited information on the structure and parameters of any model. Longitudinal studies, which are not attractive to many investigators, can provide much more understanding of the structures. The study of the variation of task characteristics with tenure and job history prior to change, carried out by Katz [17], showed clearly defined patterns of change in the subjects' responses to the instrument. Ivancevitch [16] has reported long term (18 months) growth and decay of performance changes following re-structuring of the working situation, and, in an examination of the basic postulates of expectancy theory, Mawhinney and Behling [23] hypothesised time responses for expectancy and operant models. Thiètart [34] has also proposed a model based on the Industrial Dynamics type of simulation. Such papers, however, form a very small fraction of reported studies and yet omitting the time responses from an essentially continuing process will reduce the value of studies and lead to problems of replication, low predictive power and applicability.

REQUIREMENTS OF MODELS OF MOTIVATED WORK BEHAVIOUR

The basic requirements of any model of motivated work behaviour are:

A. The model must explain:

(1) The response of individuals and the variations in responses between individuals which arise from their previous work histories and differing personal characteristics.

(2) The effects on the responses of both individuals and working groups of the social and cultural environment within which the working situation is located.

(3) The relationships governing the variation in responses over time; the dynamics of both short and long term work behaviour.

B. The following structural requirements must be satisfied:

(1) The model must be constructed from clearly defined elements which can be described uniquely, measured quantitatively or defined from quantitative input–output relationships or transfer functions.

(2) The elements must be connected in a logical structure with regard to the flows, processing and storage of internal and external information relating to the motivational system.

(3) The elements used in the model must incorporate known data on human perception and information processing.

Vol. 7
1979

STRUCTURE OF PROPOSED MODEL

The response to motivational stimulii in working situations is determined by a three level process. Each level is characterised by the type of information process involved and the time stability of the information storage.

Postulate I

(a) Motivated behaviour in working situations is a continuing series of discrete consecutive action choices.

(b) The subject's immediate actions are determined by a continuing series of short term evaluations of his perception of the situation.

(c) The information used in this evaluation process is held in short term memory which is of limited capacity.

Postulate II

(a) The information employed in the continuing evaluation process is selected from a set of Learned Constructs held in long term memory.

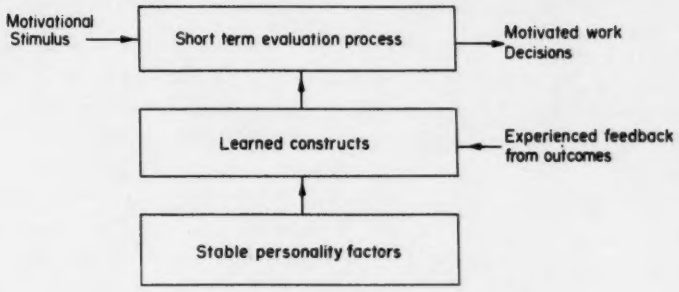


FIG. 1. *Cognitive processes in work motivation.*

(b) The set of Learned Constructs is much larger than the sub-set used in any single or short term series of evaluations.

(c) The Learned Constructs are relatively stable.

(d) These constructs are modified by feedback from experienced outcomes.

(e) The learning process will be limited by signal detection thresholds and interference from other concurrently perceived information.

(f) The learning process may follow the exponential form or may be of a dichotomous, switching, form.

(g) In the absence of feedback the information held in the set of Learned Constructs will decay by forgetting. The forgetting may be progressive for exponentially learned constructs; dichotomous constructs may switch back to the initial state.

Postulate III

(a) Each individual subject has a characteristic set of highly stable constructs, needs or drives, his 'personality', traits.

(b) These personality factors are highly stable and, under normal working conditions, changes as the result of learning are very small. High stress or traumatic experiences may induce significant changes.

(c) The Set of Stable Constructs will determine initial values of the Learned Constructs.

(d) They exercise a strong influence on the learning, forgetting and selective perception of the set of Learned Constructs.

Postulate IV

Internal motivation occurs when the subject modifies his set of learned constructs, expectations, instrumentalities, valencies,[1] by direct comparison between expected outcomes and experienced outcomes in a *single evaluation* cycle.

Postulate V

External motivation occurs when the subject updates his set of learned constructs by *re-evaluation* of the state that led to the motivated action choice.

[1]In the development of theories of motivation, some of the sets of learned constructs have been classified as expectancies, instrumentalities and valencies. There are many variations of definition and operationalisation of these terms (See Discussion (9)).

The following definitions of these classes of constructs are implied in this study and provide links for comparison with alternative theories. *Expectancy*: The subject's current set of subjective probabilities representing his perceptions and beliefs of the relationships between events:

Actions—Outcomes

Levels of performance—Outcomes

$E = 0$: no relationship.
$E = 1$: complete dependency, certainty.

Instrumentailty: The subject's current perceived set of beliefs in the relationships between different contingent outcomes:

$I = 0$: no relationship.
$I = 1$: complete dependence.

Valency: The subject's current perception of the value or utility of a specific outcome:

$V = -1$: highly unattractive outcome—maximum avoidance behaviour.
$V = 1$: indifference.
$V = +1$: highly attractive outcome—maximum seeking behaviour.

The weighted potential attraction for outcome is $0_j = E_j \times I_j \times V_j$.

BASIC MODEL

The process of response to a motivational stimulus is that the subject evaluates his perception of the situation by a valency expectancy process [35] in his short term memory. The perceptions, expectancies and valencies employed in this evaluation are selected from the current set of learned constructs held in his long term memory. The combination and summative approach of the valency-expectancy model offers the simplest paradigm for combining the subject's perception of the probabilities and attractive or unattractive valencies of possible outcomes of his action into a nett decision and choice of action.

The model differs from current Valency Expectancy (VIE) formulations in the output from the evaluation or decision-making process. The normal construct [20] is *effort* which moderated by *ability* and limited by *organisational constraints* leads to work outcomes.

It should be noted that 'effort' is used in many studies of work motivation yet there appears to be no agreement on satisfactory operational definitions in the studies surveyed.

> "It is in the measurement of effort *in situ* where we are really hurting and it would be well worth our while to start an in-depth look at the meaning and measurement of just this variable". Campbell & Pritchard [9, 11].

The construct 'effort' tends to imply a continuous 'straining for performance'. This may apply in sporting situations but rarely applies in normal repetitive production jobs. Outputs from the evaluation of a motivational situation can be: [19].

- *Choice* of behaviour pattern, from available alternatives
- *Force* of behaviour
- *Persistence* of behaviour

Detailed studies of workers carrying out repetitive tasks [6], show that the distribution of job cycle times changes from a slow broad spread to a more rapid 'normal' spread as they learn the job. With further experience the modal values tend to stabilise at a highly consistent 'natural' pace interspersed with long cycles which reduce the mean performance to the local norms or 'felt fair' pace.

Thus for the experienced worker his performance is determined by a continuing series

Vol. 7
1979

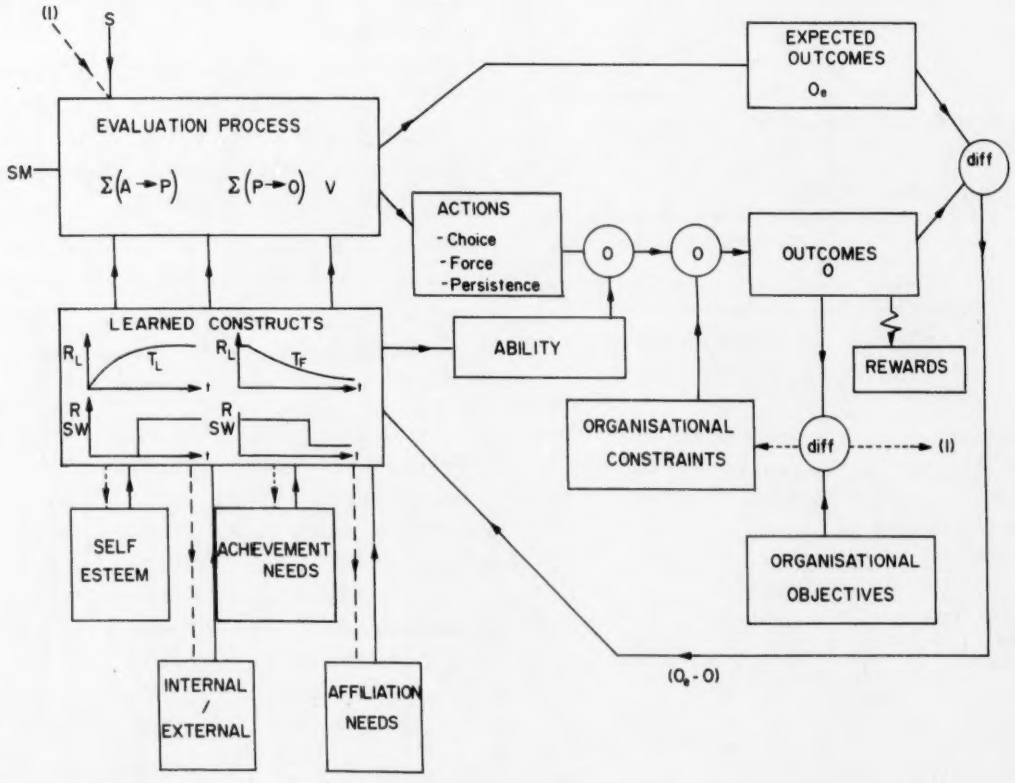


FIG. 2. *The basic motivational model.*

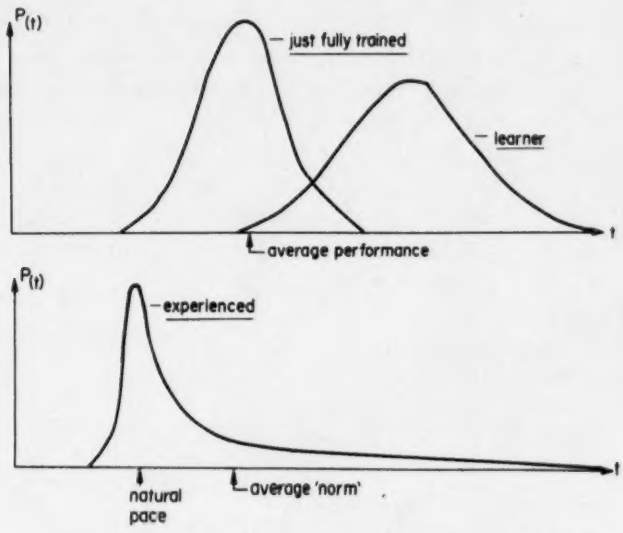


FIG. 3. *Distribution of worker performance in repetitive work.*

of choices of behaviour; on completion of one job element he can immediately select the next work related activity leading to high performance in the work task, or he can select a non-work activity which is more attractive to his current needs state and then subsequently return to another work activity. This leads to lower work performance but possibly higher satisfaction.

"Motivation is concerned with the reasons for the choice of one act or behaviour when provided with alternatives" Zedeck [38].

Objective measurements based on quantitative work outputs represent the integration of a long series of discrete decisions—to produce or, to do something else each time a choice becomes available.

A second major difference between the proposed model and the Porter–Lawler [20] type of model is in the nature of the feedback process. It is postulated that the subject, when evaluating his choice of behaviour, remembers the results of his evaluation in short term memory. If the technical situation is such that he can experience the outcomes (0) whilst the expected outcomes (0_E) leading to the actions are still remembered he can make a *direct comparison* between expectation and outcome ($0_E - 0$). This direct comparison and updating of the learned constructs leads to the most effective form of learning:

- Internal–Motivation
- Intrinsic–Motivation

If the technology of the situation results in delays between the action decision and the observable outcomes, the subject will no longer be able to make a direct comparison as the particular state of the set of constructs used will be lost from short term memory. The only process for updating the set of learned constructs is, when the outcomes are perceived, to make a *re-evaluation* of the situation leading to the action and compare this with the perceived outcome:

- External–Motivation
- Extrinsic–Motivation

There are two distinct form of change possible for learned constructs, the subject's set of beliefs, perceptions, probabilities and values relating to work. The characteristics of the change process offer a basis of classification into scalar or dichotomous sets. In the first set the state of a specific construct can exist over a range of levels and, over time, the current value changes progressively in response to the subject's perception of relevant feedback information. Such *scalar constructs* include probabilities of action, performance, outcome relationships and valencies of goals.

Other constructs may only exist in a zero or total state—binary or *dichotomous constructs*. These are characterised by time stability in either of the binary states, but from time to time may switch completely and rapidly to the alternative state.

For the scalar constructs on each repetition of the motivational situation the feedback may be zero—the perceived outcome was exactly as expected from the evaluation and the value of the construct remains unchanged. If the outcome differs from the expectation the feedback, subject to channel capacity, interfering information and discrimination thresholds, will lead to modification of the construct value, a normal exponential learning process.

For other constructs which exist only in a binary state, repeated experiences may confirm the zero or unity value or may lead to an abrupt, switching, change to the alternative value.

In the absence of feedback the values of scalar constructs may change due to forgetting, whilst the binary type are more likely to remain stable over long periods before switching back to the initial state.

The model is essentially one of *within subject motivation* and any study involves tapping the stage of the subject's set of learned constructs.

If, in his goal seeking behaviour, the subject is to remain in control of the situation, the Principle of Requisite Variety [4], states that the variety of strategies available to the controlling subject must at least equal the variety present in the situation to be controlled. This does not necessarily imply the very large lists of drives, traits or constructs proposed in earlier drive theories as combinations of a limited number of scalar constructs (constructs to which the subject assigns levels) can generate large numbers of evaluation responses. Thus five constructs each of which can be assigned three levels leads to 29,160 possible alternatives if the evaluations are sequence dependent or 243 combinations if the evaluations are sequence independent. Similarly the possibilities resulting from evaluation of three from a set of 20, three level, constructs leads to 184,680 sequence dependent alternatives or 30,780 if the nett result were sequence independant as with purely multiplicative or additive combination in the final weightings.

It is postulated in this model that the patterns observed in individual differences of response to motivational stimuli, are determined by constructs or individual personality factors which are operationally explicit and can be measured by independent reliable instruments.

Over the years very many individual traits and drives have been proposed. In the interests of economy the model employs four personality factors which have reasonably clear relationships with the perceptions employed in the proposed evaluation process and which have experimental support as influencing individual subjects responses in work situations.

Self Esteem: Behaviour—performance expectancies [13, 19, 20, 29].

Internal/External: Performance—outcome probabilities [3, 30]

Achievement Needs: Valencies of outcomes [20, 25, 27, 37]

Affiliation Needs: Valencies of outcomes [19, 20, 26]

The main feature of this set of factors acting through the set of learned constructs as moderators of the motivated responses are:

- They are characteristic to each individual.
- They are highly stable and, other than in high stress situations, are unaffected by learning or forgetting processes.
- They provide an independent method of classifying broad response patterns for individuals.

Further research may lead to modification of the proposed set or to redefinition of specific constructs. The case for retention of them in the proposed model is:

(1) They provide a link with current theories and research.

(2) As a set of classifying factors in the design and control of experiments.

(3) As the basis of taxonomies in developing field research and prescriptive application in management and work design programs.

The operation of the stable constructs with the information processing structure in individual subjects still remains speculative. The following processes are offered as basic hypotheses. The stable set of 'personal' constructs act as:

1. *Determinants of initial values* of the learned constructs held in long term memory until

Vol. 7
1979

they are modified by experienced learning from the subject's perception of the outcomes resulting from his actions.

2. *Filters in the feedback path* leading to selective perception of the outcomes and selective learning.

3. *Structural parameters* determining the learning process is in long term memory [7]. These may be selective registers in a Broadbent model or facilitators, selected routing in the rehearsal stages of an Atkinson–Schifferin cyclic selection and storage model.

4. *Determinants of rehearsal cycling* in evaluating the anticipated outcomes of possible actions leading to an initial goal seeking set of behaviours for highly salient goals. Maintenance of the selected strategies by recycling until experienced feedback cues enable the subject to refine his goals and strategies.

5. *Stable constructs as a standby* enabling the subject to maintain perceived control of a situation in the event of overloading of his information channels or, experiencing too high a level of inconsistency or delay in the feedback data to allow analysis and up-dating of the set of learned constructs.

When cues are congruent subjects use all the cues available; when they are in conflict subjects focus on one at a time, or select from an alternative set of other cues [31, 32].

THE DYNAMICS OF MOTIVATED RESPONSE

Most working situations continue over long periods of time within which change can occur. Theories of work motivation must explain the subject's short, medium and long term changes in behaviour pattern in response to both stable and changing work conditions as he reacts to stimuli from both work and non-work environments and experienced changes in his personal needs and goals. The model offers a structure of information flow, feedback, learning and forgetting processes as a framework for the analysis of dynamic response characteristics.

A subject approaches a new work situation with a set of constructs determined by his prior work and non-work experiences and his individual set of stable constructs. Once he has made decisions and adopted certain courses of action he can then obtain feedback from his personal experience of the work itself, from other people and organisational agents [14] to up-date the constructs used in his evaluation and selection of courses of action. The dynamics of this learning are determined by four main characteristics related to each subject-work situation:

1. The *Feedback Loop Cycle Time.* The total time to evaluate, act according to the evaluation, experience the outcomes and up-date the set of constructs ready for the next evaluation; these times may be short, within the span of short term memory, medium or long exceeding the short term memory span.

2. The *Time Constants* of learning and forgetting or decay. The concept of a time constant applies strictly only to constructs exhibiting an exponential form of incremental learning and forgetting [7]. There is considerable evidence that the exponential relationship between learned performance and time for both perceptual and manipulative skills provides an accurate description of human performance in work situations [6] and, as the response of a first order control system, is linked directly to the behaviour of feedback systems.

3. The *Behaviour Increment* for each cyclic repetition of his evaluation, outcome updating loop. Increments can be classified as small—many increments required to reach the saturation level of the response, intermediate and large—a high proportion of the saturation level, the whole response in the limit for a switching form of response.

4. The *Saturation Levels.* The maximum amount of behavioural change related to a specific construct in a given subject.

The response characteristics of the model can be derived from the properties of feedback control systems.

In considering the time response to motivation in working situations three main characteristic patterns may be expected.

TABLE 1.

Loop cycle time	Time constant Learning	Time constant Forgetting	Behavioural increment	Saturation level	Response type
Short	Short	Long	Small	High	A
Short	Short	Long	Large	Low	B
Long	Short	Medium	Large	High	C

The overall response of a subject can be built up from the combination of the responses of each of the set of constructs he uses in making his individual decisions in the working situation. (Curve *D*).

Simulation of the model allows the investigation of the response of the whole model when the changes in levels, of a number of constructs with differing time constants, multiple action choices and delays in feedback are too complex for simple intuitive evaluation.

RELATIONSHIPS WITH OTHER THEORIES

The proposed model is a general model of human behaviour in working situations. The underlying assumption is that each subject evaluates and responds to the work and environmental situation in an internally rational and logical manner when judged by *his* current needs and perceptions of the situation. If this were not so, there would be little hope of understanding behaviour. "Since a man's actions are guided largely by his own perception and evaluation of the situation, any incentive or disincentive must be perceived by him as such in order for it to have its fullest and most appropriate effect. A man's maximisation of interests over time is always with respect to the situation as he perceives or processes it", [24]. Motivation, the label for the determinants of goal seeking behaviour, choice, force and persistence, must accept that the range of goals sought extends far beyond those of simple economic man.

As a general model it has features in common with many other models and theories and shows considerable utility in the examination and analysis of experimental evidence relating to them. At the stable construct level there is an implied acceptance of a substantive or con-

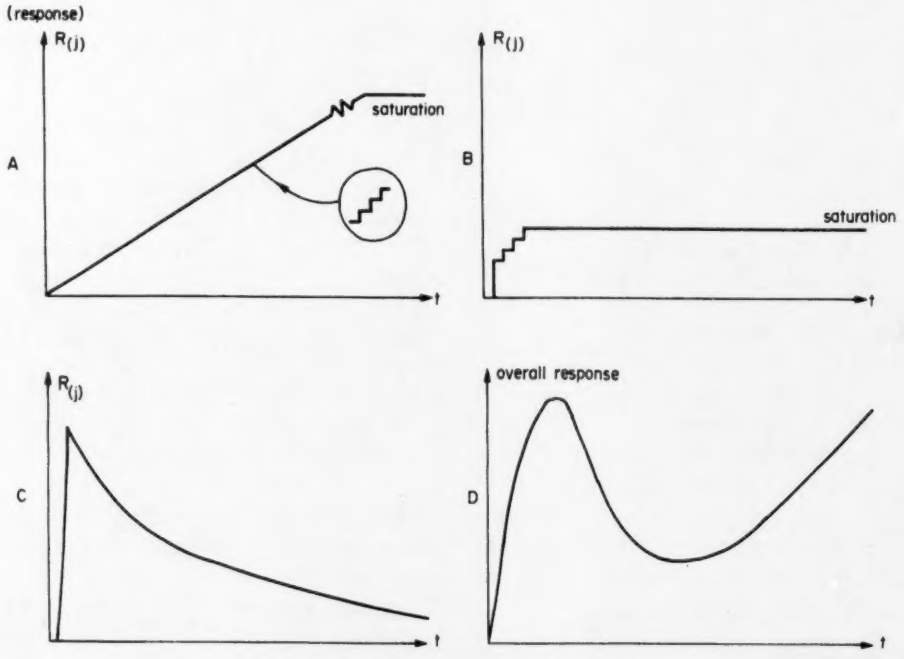


FIG. 4. *Basic forms of motivational responses.*

tent basis. The level of learned constructs has strong links with the 'Hullian' process and the operant conditioning models. The evaluation process follows the cognitive process approach using the expectancy–valency concepts.

In examining other theories within this framework, four features of the proposed model should be considered:

1. All the features of the model can contribute to the magnitude and dynamics of the subject's response to motivational factors in a continuing work situation. Any simpler model structure may show high correlations if all the other factors can be adequately controlled during observations or experiments but can only be expected to explain a small part of the total observed variance in studies of ongoing work.

2. The model is essentially one of a within subject analysis. His response depends on his current set of learned constructs which are determined by his set of stable constructs, his previous learning experience, and the subset he perceives as relevant in his evaluation processes. Though clustering of response patterns may be observed and used as a basis for classifying subjects and for experimental controls, the possible variety is very large and the predictive efficiency of such group, common factor based data will tend to be low unless the classification has a complexity matching the possible variety.

3. The model is dynamic in nature, postulating that subjects make a continuing series of sub-goal decisions in their response routes towards their major goals and that during this process both the set and values of the constructs used continually change through learning and forgetting. Thus any measurement will only reflect the state of the subject at the time of the measurement. Stability can only be attained when the subject's response variety matches that of the work and its environment and that environment is stable. In many work situations the complexity and multiple character of causal interactions inevitably lead to turbulent fields [12] and placid stability is unlikely.

4. Self report judgements of satisfaction depend on the subject's ability to integrate the continuing series of perceptions fed back from the outcomes of his activities and relate them to the motivational states that led to his actions rather than relying on a few salient incidents.

The proposed model offers a useful framework for the examination of the many theories of motivation and of experimental evidence of both successful and unsuccessful replications. In particular, the model extends the valency-expectancy paradigm [10, 20, 29, 35] in the evaluation process but differs from some recent formulations [20] in returning to Vroom's 'choice' rather than 'effort' by using discrete action choices rather than a 'pervasive effort' and, in considering the nature of the information flows and the feedback of experience from the work outcomes.

The cyclic nature of the model, a sequence of discrete behavioural choices, the result of evaluation of a limited selection from a subject's current construct set relaxes the requirement for strict transitivity in expectancy theory [5]. Intransitivities may reduce the effectiveness of the learning processes but have little effect on the long term responses.

Needs theories [2, 22], Goal Seeking [21], Two-Factor [15], Achievement motivation [24, 25], Equity Theories [1] and Operant Conditioning [11] can all be matched with parts of the proposed model. In each case the match is only partial and the remainder of the model provides a framework for analysis of the problems of testing, replication and application.

RESEARCH ON WORK MOTIVATION

The proposed model implies that work motivation is essentially a goal seeking form of behaviour. The goals and routes for their achievement are specific to each individual subject and working situation. There is an underlying assumption that this behaviour is rational in the light of the subject's current construct set, which, however, may appear irrational to an outside observer.

Performance is the aggregate result of a continuing series of action choices and can be determined objectively if the task output can be quantified and recorded.

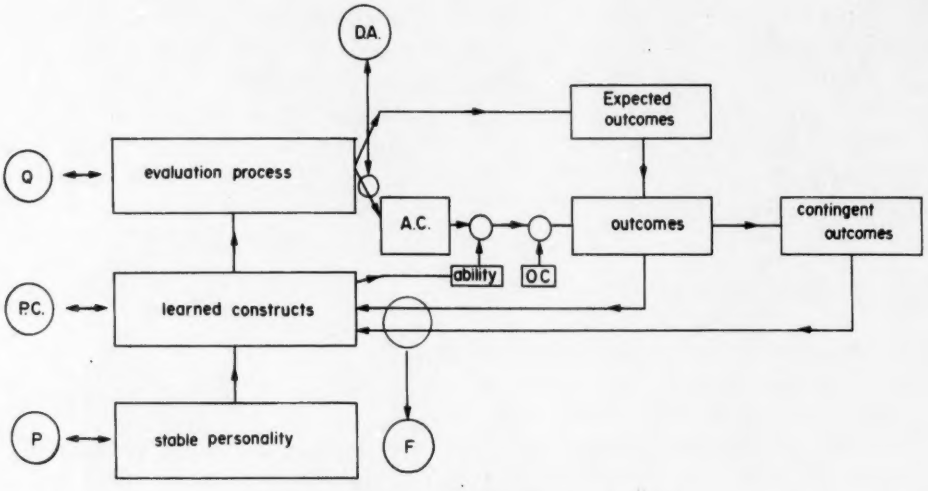


FIG. 5. *Research methods for motivational systems.*

Satisfaction results from state of congruence between his set of work related constructs and his stable personality constructs and cannot be measured directly. Satisfaction may be inferred from the subject's externally observable behaviours, quality, absenteeism, withdrawal, or by the use of self reports which are subject to the limitations of information processes.

The proposed system model centering on the flow of information in an individual's decision-making process provides a conceptual framework for the study of motivation.

One route is to measure the subject's stable personality constructs by psychometric measurements (*P*). Such measurements linked to a process theory can then be correlated with observed actions or subject's self reports. This approach omits the problems of determining the subject's learned constructs and thus may confirm general motivating factors but only explain a small part of the observed inter-subject variances.

Personal Construct theory [18] is attractive as a non-transparent technique for tapping the subject's current set of learned constructs [PC]. The problems of selection and weighting of the limited set employed in making each successive action decision is a source of variance in relating the measured constructs to the subject's work behaviour.

Traditional questionnaires tap the subject's evaluation of factors relating to his work behaviour (*Q*). An evaluation carried out in answering a question is, however, a conscious activity and the subject may use different constructs and different weightings to those in his limited set used in the 'sub-conscious' decision-making inherent in many continuing work situations.

The alternative to these approaches is to measure the information flow within the motivational system. The results of Herold & Greller's study [14] of what information subjects use in finding out how well they are doing in their jobs (*F*) provides factorial evidence for the subject's classification of feedback cues which are consistent with the structure of the proposed model. This, essentially questionnaire, methodology can be criticised as the process of thinking about feedback may be different to that of using feedback. More detailed observation of the actual cues available to workers during their work, the response to them and their perception of the use of these feedback cues may well lead to an increased insight into the parameters of the work motivated process.

An alternative approach adopted by Zedeck [38] is to present the subject with a series of decisions on combinations of varying levels of a number of working conditions and outcomes (DA). The series of decisions made by the subject on the attractiveness or otherwise of the overall task can by analysed by cluster analysis leading to measurement of the weightings used on each of the factors and thus leading to identification of relatively homogeneous subgroups. This study showed significant differ-

ences between subjective and objective weightings particularly on pay and security for some groups of subjects. As the study was directed at "would you like this job?" rather than "will you make another piece now, or would you rather take it easy and talk to Joe for a minute?". The study of choice of alternative activities within a job would appear to be a logical development of this methodology.

From the diagram it can be seen that current research approaches the study of the motivational system from many directions. Each approach however, looks at only a small part of the overall process. To date, despite the large effort in the field, the predictive success of these studies is limited.

Further advance would appear to depend on two factors:

1. A system can be specified by sets of transfer functions or operational rules for the individual elements and by information flows between the elements. Combination of measurements of states of the sets of stable and learned constructs with observation of the information flow in action choice decision making and of the form and use of feedback information from the task outcomes can provide the next step in increasing our understanding of how people react to work.

2. The motivational response is dynamic. Evidence from industrial engineering studies and longitudinal surveys or experiments indicates that a time function of 6–18 months duration is characteristic of worker's response to changes in motivational conditions over a wide range of situations.

Any thorough study of work motivation must extend over similar periods as correlations based on the relative states of factors at a given 'instant' in time when these factors may be changing at differing rates will be transient and may give little information about the underlying processes or future states.

Further development of sound quantitative theories of work motivation can only be expected from the adoption of comprehensive dynamic system models of the motivational process quantified by long term longitudinal studies of individuals and groups of workers.

REFERENCES

1. Adams JE & Rosenbaum NE (1962) The relationship of work productivity to cognitive dissonance about work inequities. *Jl Appl. Psychol.* **46**, 161–164.
2. Alderfer CP (1972) *Existence, Relatedness and Growth: Human Needs in Organizational Settings.* Free Press, New York, USA.
3. Anderson CR (1976) The relationship between locus of control, decision behaviours and performance in a stress setting. *Am. Acad. Proc.* 65–69.
4. Ashby R (1969) *Introduction to Cybernetics.* Chapman and Hall, London, UK.
5. Behling D & Stark GA (1973) The postulates of an expectancy theory. *Acad. Mgmt Jl*, **16**(3), 373–388.
6. Bevis FW & Towill F (1972) Managerial systems based on learning curve models. *Int. Jl Prod. Res.* **11**(2), 217–238.
7. Bolles RC (1975) *Learning Theory.* Holt, Reinhardt-Winston, London, UK.
8. Broadbent EE (1971) *Decision and Stress.* Academic Press, New York, USA.
9. Campbell JP & Pritchard RD (1976) Motivation theory in industrial and organizational psychology. In *Handbook of Industrial and Organizational Psychology*, (Ed. Dunnette MD), Rand McNally, Chicago, USA.
10. Dachler HP & Mobley WH (1973) Construct validity of an instrumentality–expectancy task goal model of work motivation. *Jl Appl. Psychol.* **58**(3), 397–418.
11. Dunnette MD (1976) *Handbook of Industrial and Organizational Psychology.* Rand McNally, Chicago, USA.
12. Emery FE & Trist EL (1965) The causal textures of organisational environments. *Hum. Relat.* **4**, 21–32.
13. Gavin JF (1973) Self-esteem as a moderator of the relationship between expectancies and job performance. *Jl Appl. Psychol.* **58**(3), 83–88.
14. Herold DM & Greller MM (1977) Feedback: the definition of a construct. *Acad. Mgmt Jl*, **20**(1), 142–147.
15. Hersberg F, Mausner B & Snyderman B (1959) *The Motivation to Work.* John Wiley, New York, USA.
16. Ivancevich JM (1977) Different goal setting treatments and their effects on performance and job satisfaction. *Acad. Mgmt Jl*, **20**, 406–419.
17. Katz R (1976) Career implications for job satisfaction. *Acad. Mgmt Proc.* 104–108.
18. Kelly GA (1955) *The Psychology of Personal Constructs.* Norton, New York, USA.
19. Korman AK (1970) Toward a hypothesis of work behaviour. *Jl Appl. Psychol.* **54**, 31–41.
20. Lawler EE (1971) *Pay and Organisational Effectiveness.* McGraw-Hill, New York, USA.
21. Locke EA (1976) The nature and causes of job satisfaction. In *Handbook of Industrial and Organisational Psychology* (Ed. Dunnette MD) 1293–1349. Rand McNally, Chicago, USA.
22. Maslow AH (1970) *Motivation and Personality.* Harper and Row, New York, USA.
23. Mawhinney TC & Behling D (1974) Differences in predictions of work behaviour from expectancy and operant models of individual motivation. *Acad. Mgmt Proc.* 3rd *An. Meet.*
24. McClelland DC & Winter DG (1969) *Motivating Economic Achievement.* Free Press, New York, USA.
25. McClelland DC, Atkinson JW, Clark RA & Lowell EC (1953) *The Achievement Motive.* Appleton-Century Crofts, New York, USA.
26. Nemiroff PN (1976) Task effectiveness and human fulfillment in organisations: a review and development of a conceptual contingency model. *Acad. Mgmt Rev.* 69–82.

27. OLDHAM GR, HACKMAN JR & PEARCE JL (1976) Conditions under which employer respond positively to enriched work. *Jl Appl. Psychol.* **61**(4), 393–403.
28. PINDER CC (1977) Concerning the application of human motivation theory in organisational settings. *Acad. Mgmt Rev.* **2**, 384–397.
29. PRITCHARD RD & SANDERS DS (1973) The influence of valence, instrumentality and expectancy on effort and performance. *Jl Appl. Psychol.* **57**, 55–60.
30. ROTTER JB (1966) Generalised expectancies for internal v external control of reinforcement. *Psychol. Mono.* 80–1.
31. SLOVICK P (1966) Consistency and cue utilisation in judgement. *Am. Jl Psychol.* **79**, 121–138.
32. SLOVICK P & LICHTENSTEIN S (1971) Comparison of Baysian and Regression Approaches to the study of information processing in judgement. *Org. Behav. & Hum. Perf.* **6**, 649–746.
33. STEERS RM & PORTER LW (1975) *Motivation and Work Behaviour.* McGraw-Hill, New York, USA.
34. THIETART RA (1977) Human dynamics: a new approach to human behaviour. *Omega* **5**(3), 281–291.
35. VROOM V (1964) *Work and Motivation.* John Wiley, New York, USA.
36. WARR P (1976) Theories of motivation. In *Personal Goals and Work Design* (Ed. WARR P) 141–167. John Wiley, New York, USA.
37. WEINER B (1972) *Theories of Motivation: From Mechanism to Cognition.* Markham, Chicago, USA.
38. ZEDECK S (1977) An information processing model and approach to the study of motivation. *Org. Behav. & Hum. Perf.* **18**, 47–77.

ADDRESS FOR CORRESPONDENCE: *Martin Walbank, Esq, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK.*

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 4. pp. 309 to 320
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0801-0309$02.00/0

Aristotle's Fallacy: A Hypergame in the Oil Shipping Business . . . .

MO GIESEN

PG BENNETT*

University of Sussex, UK

(*Received October* 1978: *in revised form January* 1979)

In this paper, an attempt is made to illustrate and validate some predictions of the 'hypergame' model of decision-making, a generalisation of the conventional game model as first described by von Neumann and Morgenstern. The hypergame concept, which can take players' differing perceptions, misperceptions and information conditions into account, is formally defined and two simple criteria of stability and choice are given. The core of the article consists of a case study involving decision-making under misperceptions. One result of the analysis is to show that decisions which at first sight appear 'stupid' and remain 'irrational' in a conventional game model, are indeed 'rational' if the players' perceptual limitations and their differing information conditions are considered. The conclusion is that the hypergame can be a useful extension of the traditional theory of games, and as such is one step on the way toward conceptually more complex and 'realistic' model building.

INTRODUCTION

ONE OF THE major advances in mathematical decision-theory was the Theory of Games as first presented by von Neumann and Morgenstern [18]. This dropped some of the assumptions of the then prevailing economic theory, and instead considered decisions made by competing and cooperating individuals with logical rigour. However, the axioms which made this rigour possible sometimes appear too stringent and the models are thus not always relevant to 'real world' decisions. Perhaps it was the emphasis on the *solvable* model which sometimes conflicted with the original purpose of the theory, to give explanation and guidance in complex decision problems.

One of the assumptions of a traditional game model is that of the 'well informed player', who knows the full range of his own and his opponents' strategies and preferences or utilities, and has the same tools and abilities of analysis. It is quite obvious that in many decision problems such complete information of all players cannot be assumed. This led to the introduction of the hypergame model which relaxes the assumption of complete information, and instead relies on players' perceptions as the basis for analysis. The hypergame model is by no means the first or the only step in this direction, other writers have explored a number of ways of extending game theory (notably [10]). An advantage of the hypergame seems to be that it is a comparatively straightforward generalisation of the traditional model, and that traditional choice and stability criteria can be applied with only minor modifications.

The following case study is intended to be an example of the application of the hypergame model to a decision problem, albeit a retrospective application. It is thus an element in a series of studies carried out to validate, or at least show the range of applicability of

* Work supported by SRC grant GR/A/16159.

the hypergame model. Another study in the context of military decision-making has already been presented [4], and in cooperation with a major company the model is currently used to examine the difficulties presently experienced by the shipping industry [11].

In addition to the features already exhibited in other case studies, the hypergame model presented here high-lights the importance of cross-game information, an issue arising from the perceptual framework employed. It seems that in a large number of conflict problems the parties' ignorance or knowledge about each others' differing perceptions can be decisive, and thus certainly needs to be incorporated into a model of conflict situations under conditions of incomplete information.

THE THEORY OF HYPERGAMES

The formal definition of a hypergame in its most general form and of the stability and choice criteria which will be used in this case study is based largely on that given in [1]. Since the following is an application of hypergames and does not present any further developments of the formal theory, it was considered sufficient to give the definition of a hypergame and the criteria in a more informal fashion. It was also assumed that the reader would be familiar with at least some of the game-theoretic concepts explained in most standard textbooks (e.g. [13]).

The hypergame can be considered as a stack of games, one for each player's perception of the situation. More formally, an n-person hypergame in normal form consists of,

(1) a set of n players;

(2) for each player, n sets of strategies, i.e. those strategies that he perceives each player—including himself—to have;

(3) for each player, n preference orderings over the set of outcomes generated by the above strategies, expressing the preferences he perceives himself and the other players to have.

An outcome of the hypergame is reached when each player chooses one of the strategies he perceives himself to have. Since players' perceptions of the available strategies may be different, it is possible that an outcome of the hypergame does not appear in a given player's game, thus introducing the possibility of a strategic surprise.

Naturally, each player makes his choice in full knowledge of his 'own' game, and in such a way as to obtain an outcome as highly preferred by him as possible; however, various possibilities of *cross-game information* arise. The first is that of completely restricted information. That is, no player has any knowledge about his opponents' differing perceptions, so that each sees only his 'own' game. The opposite of this is a hypergame with complete cross-game information in which all players know each others' games. Between these two extremes lie hypergames with partially restricted information in which some (but not all) players have knowledge of at least one of their opponents' games. These information conditions are of considerable importance when hypergame stability is considered. The case study which follows shows that this is particularly so if stability criteria are employed as criteria of choice.

The numerous ways in which differences in players' perceptions may arise are examined in more detail in [2], but it seems worth mentioning here that they can be divided into two categories. The first includes all disagreements about the sets of strategies available, i.e. whether certain strategies are physically possible or permissible, or when the limitations of the mental resources of some players prevent them from examining all possible strategies. The second category comprises disagreements about preference orderings. These may happen through a misperception of a player's value system or through disagreement about the factual aspects of the case, i.e. the real physical consequences of a formal outcome. It should be noted that misperceptions can also be divided into ones which are accidental and those which are deliberately induced. (On the subject of deception in terms of conventional 2×2 games see [7]).

STABILITY AND CHOICE IN HYPERGAMES

Here, stability and choice criteria will be considered only in the context of 2-person hypergames. However, the extension to more players seems straightforward. Mainly because

Vol. 7
1979

of their simplicity, only two criteria of choice and two stability criteria will be defined, although others could equally well be adapted to hypergames. Before going into any detailed explanations it should be pointed out that according to these particular criteria only outcomes appearing in both players' games are considered to be possible stable outcomes of the hypergame.

The first criterion to be considered is that of *α-stability* which defines an outcome as stable if no 'unilateral improvements' are possible. This is an extension of the game-theoretic concept of an equilibrium outcome [15]. More formally this can be expressed as:

(1) An outcome is *α-optimal* for a player if and only if he cannot obtain a more preferred outcome by changing his strategy choice, while the other's choice is held constant;

(2) An outcome is *α-stable in the hypergame* if and only if it is an α-optimal outcome for each player in his respective game.

As a special case of this we introduce the notion of *weak α-stability*. For a player an α-optimal outcome is weakly α-optimal if there *is* an outcome obtainable by a 'unilateral' strategy change (as described under condition (1)), ranked *equal* to it in the player's preference ordering. An outcome of the hypergame is weakly α-stable if it is α-stable but only weakly α-optimal for one or both players. Due to the 'unilateral' character of α-optimality, an outcome of the hypergame is α-stable regardless of the misperceptions the players might have about their opponent's preferences. This is not the case for the second criterion, which is based on a different line of reasoning. Roughly speaking, an outcome is considered to be β-optimal by a player if a strategy change, followed by his opponent's 'best reply', does not bring any improvement for the first player. Again, an outcome is stable in the hypergame if it is β-optimal for both players in their respective views. Since this criterion is dependent on one's perception of the other player's preferences, it is also dependent on cross-game information. The other's 'best reply' can be evaluated in *his* game (if one has cross-game information), but it is necessarily evaluated in *one's own* game (if one has restricted information). More formally this can be expressed as:

(1a) For a player with *restricted information* an outcome is *β-optimal* if and only if there is no outcome in his own game ranked higher in his own preference ordering and α-optimal for the other player.

(1b) For a player with *cross-game information* an outcome is *β-optimal* if and only if there is no outcome ranked higher in his own preference ordering and α-optimal for the other player in that player's game.

(2) An outcome is β-stable in the hypergame (under the specified information conditions) if and only if it is β-optimal for each player in his respective game.

As before, we introduce the idea of *weak β-stability* as a special case. An outcome, β-optimal for a player, is weakly so if there is an outcome as described in the appropriate criterion (1a) or (1b) ranked equal to it in the player's preference ordering. A β-stable outcome is weakly β-stable in the hypergame if it is weakly β-optimal for one or both players.

Without going into a detailed discussion of the notion of stability in general, let it suffice to say that it is considerably less ambitious than the concept of the 'solution' of a game, (*cf.* [18], especially p. 31). It does not impose a norm of 'rational behaviour' but simply states that stable outcomes, once reached, will persist as long as the players adhere to the line of reasoning given in the criterion.

More normative in character is the idea of a criterion of choice. Here, we will make use of the well-known game-theoretic notion of a *dominant strategy*. A strategy for a given player is dominant if, whatever the other player's strategy choice is, choosing it yields an outcome at least as highly preferred as could have been obtained by choosing any other. A player may have several such strategies; however, we shall not meet any cases of this type here. This being so, we can simplify the formulation of our criteria considerably by assuming that no player has more than one dominant strategy. We can then define *two criteria of choice*—applicable both to games and hypergames. The first of these specifies that a player having a dominant strategy should choose it. The second criterion

states that a player who perceives the other to have a dominant strategy should assume that he will use it and choose the strategy that is his 'best response' to it. Clearly these two criteria will be compatible. Bearing in mind our previous comments about 'solutions' in games, it should be emphasised that we are not claiming that these axioms provide one with the 'right answer' even in misperception-free games where dominant strategies exist—the 'Prisoner's Dilemma' being a case in point. Nevertheless, it is interesting to trace the application of the criteria to the cases we shall be dealing with.

THE CASE HISTORY

The following chapter will give the essential background for the hypergame analysis of one of the crucial decision situations in an incident which in the year 1954 nearly led to the bankruptcy of Aristotle Onassis, the Greek tanker fleet owner. Besides Onassis, the main parties involved in the controversy were Stavros Niarchos, Onassis' main competitor, ARAMCO, the American owned giant oil trading and refining company, and King Saud of Saudi Arabia.

The details for this story were taken from several newspaper reports (mainly an article in the *Sunday Times* of October 9th 1977—hereafter referred to as [*ST*][1]), several accounts of the exploitation of oil resources in the Middle East [14, 17] and a biography of Onassis [6]. While the historical accuracy and completeness of some of our sources could be doubted, the main concern of this case study is with the interpretation of the available material. However, we think that our presentation is the most plausible of a number of alternatives emerging from this set of (often confused and occasionally contradictory) accounts.

The series of events of interest here began in the year 1953. By this time Aristotle Onassis had accumulated one of the world's largest fortunes in shipping under contracts with all major American oil companies. But while these contracts were highly profitable, they also made Onassis completely dependent on the 'good will' of the American companies, a dependency from which Onassis would rather have freed himself. He also knew that the oil producing countries were similarly dependent on the oil companies, in so far as they had given exploitation and transportation monopolies to companies such as ARAMCO or the British AIOC. Some years ago, in 1950, Persia had tried to become independent of the AIOC, with which it had such a contract, but the oil companies imposed a shipping embargo on Persian oil and she had no choice but to sign a new concession. This led Onassis to believe that a contract between an independent shipowner, such as himself, and an oil producing country would not only solve his own problem but would also be appreciated from the producer's side.

With this in view, Onassis had allegedly approached several large oil producing countries (Iran, Iraq, Venezuela and Kuwait according to [6, p. 90]), but his most promising candidate was Saudi Arabia. This was then under the rule of King Saud, who derived huge revenues from a concession given to ARAMCO in 1933; but despite this, Saud was in constant financial crisis (mainly due to his rather excessive lifestyle), a fact which Onassis reckoned would make him more amenable to his plans. With this in mind he carefully approached him through a middleman, Spyridon Catapodis. This led to the secret signing of a draft contract in December 1953, and the main agreement was signed by Saud, Onassis and Catapodis on January 20, 1954. In one source this contract is described as follows:

> "Onassis was to set up the Saudi Arabia Maritime Co. (SAMCO) to run the ships he put into trade. It would have its headquarters somewhere in Arabia. The Arabian government was to hold a controlling interest in the company and Onassis was to receive a substantial share of all profits made. When ARAMCO's ships eventually retired, Onassis' SAMCO tankers would take over, giving him an eventual monopoly on the transportation of most of the oil produced in the world, a business manipulation that would have meant hundreds of millions, if not billions of dollars of profit to Onassis over a relatively short period.
>
> One of the important stipulations of the contract was that the estimated 40 million tons of oil to be shipped each year was to be carried at a rate higher than the normal market price. In return, Onassis agreed to pay a royalty of 21 cents on each ton of oil to the government for the exclusive right to carry the oil".
>
> [6, pp. 89, 90]

Although the signing of the agreement was announced publicly, its exact terms were to be

[1] This article contains major extracts from a biography of Onassis published as [8].

Vol. 7
1979

held secret. After this successful start, Onassis and Saud were surprised and dismayed when, after having returned to his villa at the Côte d'Azur, Catapodis leaked information about the agreement to his friends and subsequently to newspapers (Onassis' decision to include Catapodis in the deal and the latter's leak of information can also be analysed in terms of hypergames and this was done in [9]). Obviously ARAMCO also took notice of these newspaper reports and, by exerting pressure on King Saud, forced him and Onassis into lengthy negotiations about modifications to the contract. After strong representations by ARAMCO and seven Western governments the contract was amended on April 7, 1954. It was promised that ARAMCO could keep a certain share of the transportation business and SAMCO would charge current market rates. But ARAMCO and the US Government were not satisfied.

> "Considerable pressure was brought on the Saudis through orthodox diplomatic channels. The U.S. Ambassador in Jidda, Edward Wadsworth, under direction from John Foster Dulles and the State Department, persuaded King Saud to meet with ARAMCO and by August the King told Wadsworth that he wished 'to get out from under' the Onassis deal. There remained however, the question of the Arab's 'amour propre'. The King had to be given a satisfactory pretext for breaking the agreement." [*ST*]

News about the contract between Onassis and Saud had also alarmed Stavros Niarchos, who had then hired a former CIA agent, Robert Maheu, and given him the task of undermining the deal. When Maheu heard about Saud's admission that he would like to abandon the deal, he set out to find the excuse which the King was looking for. It was Maheu's luck that by this time Onassis and Catapodis were not on speaking terms anymore. Onassis blamed Catapodis for breaking the secrecy surrounding the contract and Catapodis felt 'cheated' by the renegotiations between Onassis, Saud and ARAMCO in which he was not taking part. The animosity between Onassis and Catapodis led the latter to sign a statement for Maheu, alleging that an income tax exemption for Onassis' company had been obtained by a scheme of bribes amounting to £500,000 and paid to high Saudi government officials. When Maheu presented this statement to King Saud, he was told that the King was personally grateful for the information and had decided to 'have done with Onassis' [*ST*]. Immediately after this Saud invited Onassis for an audience to Riyadh, Saudi Arabia's capital.

At this time Saud and Onassis were still apparently trying to modify the contract to accommodate ARAMCO's demands. Thus Onassis was completely taken by surprise when at the audience King Saud confronted him with Catapodis' statement, demanding an answer to these charges and insisting that the contract should now be totally renegotiated. Onassis defended his position well, pointing out the weaknesses of Catapodis' evidence (one of the most important pieces of evidence could not be produced since according to Catapodis it had been signed in vanishing ink) and refused to renegotiate the contract. He stated in a press conference in Monte Carlo that there was 'no question of revising the whole agreement' (*The Times* of Nov. 9, 1954). Straight away, Saud relayed this information to ARAMCO and suggested submitting the contract for arbitration. The case was placed before a Swiss judge on Feb. 23, 1955. The arbitration tribunal decided to invalidate the contract by denying the Saudi Arabian government the priority it had taken for granted. This was based on the consideration that,

> "Since ARAMCO's concession is a concession for the development of national wealth, which is contractual in character and not a public service concession, the rights and obligations of the concessionary company are in the nature of acquired rights..."
>
> [17, p. 62]

The distinction between a 'concession for the development of national wealth' and a 'public service concession' is perhaps rather artificial and the true reasons for the judgement are probably spelt out more clearly in the following passage:

> "The granting state cannot compel the company to give the right of using these installations to any third parties *who, by their intervention, can but disorganise the international system* of export and sale which ARAMCO has adopted... [17, p. 62] (our emphasis)

This demonstrates well that by letting the contract go to arbitration, Saud effectively agreed to support ARAMCO's position of absolute dominance in the trade with Saudi Arabian oil, without giving the impression of breaking the agreement made with Onassis.

Saud had been given precisely the excuse which he had needed. The consequences of this decision in court only began to dawn on Onassis some time later when he found that oil companies now refused to give him any charter contracts; Onassis stated later that,

> "I made one great mistake. I never thought that in our world emotions could override all business reason. For almost a year, I had reports that Saudi Aradia was looking for someone to provide it with a fleet to carry its surplus oil.... Finally, I became convinced that somebody was about to make the deal and rather than let it go to competitors I went for it myself. And landed in the biggest mess of my life."
>
> [6, p. 95]

For Onassis the boycott implied effectively a withdrawal of his credits, without which he could not operate his fleet. Driven to the edge of bankruptcy only the 'lucky break' of the Suez Crisis with its increased demand for tanker capacity saved Onassis from complete financial ruin.

A HYPERGAME MODEL OF "ARISTOTLE'S FALLACY"

Having obtained a general case history, our attention will now be focussed on the final decision before arbitration in which Onassis, confronted by Saud with Catapodis' statement, refused to renegotiate the contract, thereby leading to submission to the arbitration tribunal and the subsequent ruling for ARAMCO.

In 1954, as explained above, King Saud already had one main aim and that was to abandon the whole deal with Onassis. But this had to be accomplished without any great loss of 'face' or money. These conditions implied that either Onassis would have to agree to have the contract declared invalid or it would have to be invalidated by some other means, for example by a 'respectable' court ruling. It seemed unlikely that Onassis would be willing to 'drop' the contract on his own account. The huge financial benefit which Onassis stood to make from the deal, coupled with the advantage of becoming independent of the oil companies, gave him an incentive to stand by the contract—even if it had to be substantially revised.

> "Onassis continued to push for the deal's acceptance because of a lag in the oil business and because the other oil companies he was dealing with were not committing themselves for long periods.... With the Saudi Arabian contract safely in his pocket, it appeared that Onassis had, in effect, defeated the oil companies."
>
> [6, p. 91]

For Saud and ARAMCO this meant that if the contract was to be invalidated at all, it would have to be against Onassis' wishes.

It was then that Robert Maheu went to Riyadh and offered King Saud Catapodis' statement and accompanying pieces of evidence. But this evidence was rather weak, and since Catapodis was well known to be a fairly unreliable character, Onassis could have countered these allegations convincingly. This ruled out simply relying on them as a justification for letting Onassis down and allowing the contract to be submitted for arbitration.

This consideration led Saud (perhaps inspired by Maheu) to form a slightly different plan. Although Catapodis' evidence was weak, it was strong enough for Saud to justify renegotiations. Onassis, on the other hand, could be expected to deny Catapodis' accusations and thus to refuse to renegotiate on these grounds. King Saud could then let the contract go to arbitration without any great loss of 'face'. He would not appear to betray Onassis but would simply draw the consequences from Onassis' 'inflexible and unreasonable' behaviour. This reasoning is also shown in the following quotation:

> "..., King Saud's patience with the deal was nearly exhausted. He privately informed ARAMCO and the US embassy that if Onassis could be shown to refuse a reasonable renegotiation of the agreement he would gladly let it go to arbitration. This was exactly what they wanted to hear. Within two months the deal was effectively off." [*ST*]

It was clearly the crux of the matter that Onassis would be unable to see that his refusal to renegotiate, albeit for good reasons of his own, would lead to an opportunity for Saud to let the contract go to arbitration without damaging his reputation.

This whole problem can now be formulated in terms of perceptual games, with the corresponding strategies and preference orderings. The range of options which Saud perceives for himself can be condensed into three possible strategies: The first is referred to as the 'soft' strategy and implies that Saud would confront Onassis with Catapodis' statement but be con-

Vol. 7
1979

tent with only minor modifications to the contract. In Saud's second strategy, termed 'hard', he would insist on complete renegotiations of the contract. Finally, Saud's third strategy, 'break', implies that he does not honour the agreement and either allows ARAMCO to submit the contract for arbitration or waits for Onassis to take action. In response, *Saud sees Onassis* as having two options. The first one is a 'soft' strategy requiring Onassis to give in to most of Sauds' demands, while the second, his 'hard' strategy, entails his virtual refusal to renegotiate.

Our sources indicate that it is unlikely that Onassis ever seriously considered the possibility that Saud might break the contract unilaterally or give his consent to ARAMCO's court actions; on two occasions in statements to the press (*The Times* of November 9, 1954 and of January 20, 1955) Onassis emphasised that he had (or at least he thought he had) 'complete support' from King Saud. Even after the contract had been submitted for arbitration, Onassis was still under the impression that Saud was on 'his side'. This misperception was further supported by the fact that Onassis, Saud and ARAMCO were still negotiating minor changes apparently with the aim of trying to satisfy ARAMCO. Such negotiations were carried out even while the arbitration was going on and thus probably represented not only a 'cover-up' for King Saud, but also a safety mechanism for Saud and ARAMCO, in case the tribunal declared the deal to be legitimate. Apart from the missing 'break' strategy Onassis' perception of his and Saud's strategies coincides with the latter's; Onassis, therefore, perceives a 2-person 2 × 2 game while Saud's perceptual game has a 2 × 3 matrix.

However, the most important differences between the two games are *the expected factual consequences of the formal outcomes*, and hence the preference orderings. This can be shown in form of a table such as Fig. 1.

Using this table, and some assumptions about the players' perceptions of their own and each others' aims, it is not difficult to construct preference orderings for the outcomes of the two perceptual games.

Let us look first at *Onassis' game*. For himself Onassis would clearly most prefer his own 'victory' $\langle h, s\rangle$, in which the contract only undergoes minor modifications. The compromise outcomes $\langle s, s\rangle$ and $\langle h, h\rangle$ occupy middle positions, and the 'victory' for Saud $\langle s, h\rangle$, is ranked lowest. Since in his opinion Saud is basically still willing to support the agreement but is trying to obtain more favourable terms (why else would he put such emphasis on Catapodis' poorly supported statement?), while he wants to minimise such changes, Onassis perceives Saud's preferences as exactly opposed to his own. We thus have the game on the Left Hand Side of Fig. 2.

But in fact Saud's aims are not as Onassis sees them. They differ not only in that Saud has an 'extra' strategy (it could actually be shown that this would not influence Onassis' strategy choice anyway), but also because *he has different aims in mind and can see a way of reaching them that Onassis cannot see.* Let us consider *Saud's game.* Since he would like to see the contract invalidated with minimum fuss, his most preferred result is the $\langle h,h\rangle$ outcome, which in his opinion would lead to a court ruling against the contract. It would also avoid giving an impression of collaboration with ARAMCO, which would have damaged Saud's business reputation considerably. Saud's least preferred outcome is obviously Onassis' 'victory', $\langle h, s\rangle$, though from his point of view the outcome $\langle s, s\rangle$ is only slightly better, since it does not satisfy his aim of abandoning the contract either.

The intermediate positions in the ranking are occupied by the pair of equivalent outcomes ($\langle s, b\rangle$ and $\langle h, b\rangle$), arising if Saud uses his 'break' strategy (Onassis' choice in this case is irrelevant), and the outcome $\langle s, h\rangle$—Saud's 'victory' in renegotiations. Without further assumptions, the relative ordering of these is uncertain. However, analysis of the hypergame will show that in most respects, the model is insensitive to changes in these intermediate rankings. For the present we shall assume that because of the strong pressures brought to bear by ARAMCO, Saud by now prefers abandoning the contract and the subsequent 'loss of face' to even substantial modifications so that $\langle s, b\rangle$ and $\langle h, b\rangle$ are ranked *above* $\langle s, h\rangle$. This seems a plausible enough suposition, bearing in mind the extreme lengths to which the established oil concerns were known to be prepared to go in defence of their positions.

In Saud's view, Onassis' preferences are opposed to his own, except that he sees the

Vol. 7
1979

Onassis' choice	Saud's choice	Formal outcome	Consequences as perceived by Onassis	Consequences as perceived by Saud
'soft'	'soft'	<s,s>	'compromise' some modifications to the contract	as for Onassis
'soft'	'hard'	<s,h>	"Saud's victory" complete renegotiations	as for Onassis
'hard'	'soft'	<h,s>	"Onassis' victory" minor modifications to the contract	as for Onassis
'hard'	'hard'	<h,h>	'compromise' some modifications to the contract	'court case' Onassis appearing unreasonable
'soft' or 'hard'	'break'	<s,b> or <h,b>	not seen by Onassis	'court case' with some damage to Saud's reputation

FIG. 1. *Outcomes and their perceived consequences.*

latter's preference for the $\langle s, s\rangle$ outcome essentially equal to those obtained when he uses his 'break' strategy. Perhaps here the case for slightly differing preferences could be argued; however, the point of real importance is that according to Saud's evaluation of the situation, the result least favourable for Onassis is certainly the $\langle h, h\rangle$ outcome opening the way to court proceedings but preserving Saud's reputation. Saud's game is thus shown on the right hand side of Fig. 2.

We now analyse the resulting hypergame. To do so we must take into account the information conditions prevailing. As we have seen, Onassis was quite unaware of Saud's real aims when called to Riyadh. The press statement asserting that he had 'complete support' from Saud [*The Times* of Jan. 20, 1955] indicates not only that he was mistaken about Saud's real intentions, but also that he was probably trying to influence ARAMCO and thus did not know about the information which ARAMCO had received from Saud through Wadsworth and Maheu. Thus he certainly did not have any cross-game information and regarded his perceptual game as *the* game being played. On the other hand, it seems that Saud was quite aware of Onassis' perception of the situation—though he obviously regarded it as a misperception. Not only did he have access to Onassis' press statements, but he even helped to create his misperceptions by disguising his true aims and mentioning the possibility of renegotiations. We are thus led to suspect that Saud had complete cross-game information.

This suspicion is strongly reinforced by analysis of the hypergame of Fig. 2. We shall see that, from Saud's point of view, cross-game information is of crucial importance in determining one's 'rational' strategy choice. To show this, we analyse the system both under completely restricted information conditions and with cross-game information for Saud. *In Onassis' game* each player has a dominant strategy, 'hard': the resulting outcome $\langle h, h\rangle$ is α-stable and β-stable (the outcome $\langle s, s\rangle$ is weakly β-stable, but it is not α-stable). *In Saud's game*, neither player has a dominant strategy. The outcome $\langle s, b\rangle$ is weakly α-stable, and is β-stable, while $\langle h, b\rangle$ is not α-stable but

Vol. 7
1979

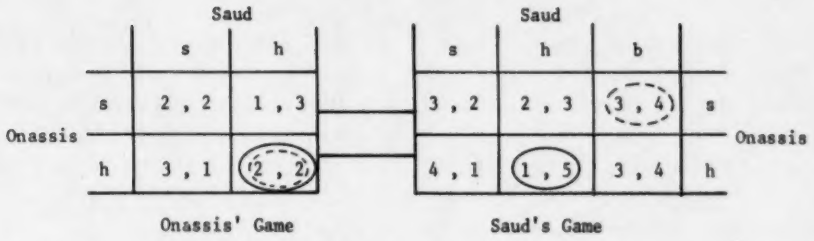


FIG. 2. *Onassis–Saud hypergame (assuming Saud prefers the* $\langle s, b \rangle$ *and* $\langle h, b \rangle$ *outcomes to the* $\langle s, h \rangle$ *outcome)*

Figures in cells denote preferences for each player, with more preferred outcomes assigned higher numbers, and 'row' player's preferences shown first. Stable outcomes of the *hypergame* are encircled by unbroken lines; outcomes stable in the individual games—omitting weakly stable outcomes—are encircled by broken lines.

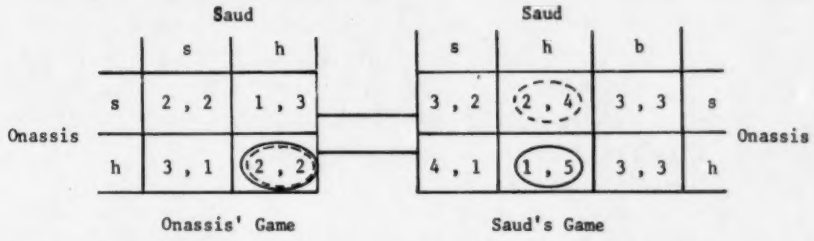


FIG. 3. *Onassis–Saud hypergame (assuming Saud prefers the* $\langle s, h \rangle$ *outcome to the* $\langle s, b \rangle$ *and the* $\langle h, b \rangle$ *outcomes)*

is weakly β-stable. No other stable outcomes exist in either of the two games.

In the hypergame with completely restricted information, let us first see what happens when we use the above considerations as criteria of choice for each player in a 'one shot' decision problem. We find the following to be a likely possibility: Onassis chooses 'hard' (expecting Saud to choose 'hard' also), while Saud plays 'break' (expecting Onassis to play 'soft'). Thus, in the hypergame, we arrive at the outcome $\langle h, b \rangle$; that is Onassis adopts a 'tough' bargaining stance and Saud unilaterally abandons the agreement. This result is quite favourable for Saud, but not as good as the outcome $\langle h, h \rangle$ obtained in 'real life'. What of stability in the hypergame? Because the 'break' strategy appears only in one of the games, our criteria do not apply to the outcome $\langle h, b \rangle$. But it is noteworthy that this outcome (and also $\langle s, b \rangle$) is β-optimal for Saud. $\langle h, h \rangle$, however, is both α- and β-stable in the hypergame. Thus, even with completely restricted information, we find that the outcome that actually occurred is the only one stable in the hypergame. The fact that it is stable, however, is no guarantee that it would be reached: in particular, under restricted information conditions *there is no clear rationale for Saud to choose his 'hard' rather than his 'break' strategy.*

In the hypergame with cross-game information for Saud, this shortcoming is convincingly remedied. From Onassis' game, Saud can see that his opponent perceives himself to have 'hard' as a dominant strategy. It is now quite clear that Saud's best tactic is to anticipate this choice and to adopt 'hard' himself in response. Thus, each player chooses 'hard', expecting the other to do so as well. The formal outcome resulting is that anticipated by both players, but the factual consequences of this are seen quite differently. Under these information conditions $\langle h, h \rangle$ is again the only outcome stable in the hypergame, it is both α- and β-stable. Now, the outcomes $\langle s, b \rangle$ and $\langle h, b \rangle$ are no longer β-optimal for Saud: because he can predict that Onassis will choose 'hard' his unequivocal 'best policy' is to choose 'hard' himself. The introduction of cross-game information for Saud thus provides an ample logical basis for his actions, and gives us a compelling model of the interaction. (We may put the argument another way: though the information conditions do not affect the hypergame stability as such in this case, the choice of the 'hard' strategy by Saud can only be regarded as 'rational'

if he is able to see that the $\langle h, h \rangle$ outcome is in fact stable. Under restricted information conditions he could not have done so: since this outcome is not stable in his individual game, it would have been more reasonable to choose his safer 'break' strategy.) In this respect this case is strikingly similar to that of the fall of France discussed in [3] and [4].

AN ALTERNATIVE MODEL

As was stated above, we have no direct evidence that Saud would have preferred the outcomes $\langle s, b \rangle$ and $\langle h, b \rangle$ to the outcome $\langle s, h \rangle$. We should therefore investigate the case in which this ordering was reversed: that is, in which Saud prefers to keep the contract provided he can obtain considerably better terms, rather than to break with Onassis unilaterally. If all other assumptions remain unchanged, we then have the hypergame represented in Fig. 3.

Once again it can be shown that the only stable outcome of the hypergame is $\langle h, h \rangle$ which is both α- and β-stable. Now, however, *'hard' is a dominant strategy for Saud*: it is clearly his best choice regardless of cross-game information. Even if Saud plays 'soft' instead of the expected 'hard', Saud wins substantial concessions, and so obtains what is now his second-best outcome. It can thus be seen that this alteration to the model does not affect the stability in the hypergame, though it does make Saud's choice 'easier'. It should be emphasised, however, that we regard the first model as the more plausible.

CONCLUSIONS

One of the most immediate questions arising from the case history and the decision analysis is how Onassis could have fallen into this rather simple trap. Surely an experienced and certainly clever businessman would expect such schemes—not least because he would have been just as capable as his opponent of devising them, and had probably on occasion even employed them himself.

It is hoped that the hypergame model has shown that Onassis was not acting 'stupidly' and certainly not 'irrationally' when he chose his 'hard' strategy and refused to renegotiate. *In terms of his perception of his own and of Saud's aims, Onassis made a perfectly sensible decision.* This is clearly shown by the fact that he chooses the dominant strategy in his perceptual part of the hypergame. The model demonstrates that it was the limitation of his perceptual framework that caused him to make a decision which in the eyes of the well informed observer looks 'stupid'.

There are very good reasons for the rather poor quality of this framework. Perhaps the most important one is that Onassis was in a state of crisis when he had to make his decision. Onassis had been called to Riyadh for no apparent reason, and he had to negotiate in an unfamiliar and—as described by Robert Maheu—rather imposing if not frightening environment. This created the sort of conditions which introduce strong perceptual limitations.

> "Alas, it is precisely under these conditions that expanding one's world view becomes most difficult from a psychological point of view."
>
> [5, p. 41]

For Onassis this meant that he concentrated on invalidating Catapodis' statement. It was agreed by witnesses later on that Onassis gave a very good account of himself in Saud's audience [*ST*], and by drawing attention to Catapodis' dubious character defended his position very well. *The fact that his defence was not only anticipated but even aimed at by Saud was outside the focus of Onassis' attention.*

The other factor limiting Onassis' perception was the realisation that his last hope of saving the contract was about to be crushed. But it was precisely at this time that he really needed the agreement.

> "Onassis' multiple ship construction was fast being completed, and he would have to provide business for them in a hurry or suffer a tremendous loss as some of his new ships went idle. He had 375,000 tons on order, with $35 million yet to be paid on the ships. Within a short while those launchings, plus the ships he owned, would have given him 500,000 idle tons in a less than dynamic market."
>
> [6, p. 91]

All this made Onassis extremely anxious to secure success for the agreement; renegotiations, as demanded by Saud, did not seem a step in the right direction. But while worrying about the detrimental effects which renegotiations would have had on his profits, Onassis left the possibility of King Saud bringing the

Vol. 7
1979

case to court, or encouraging ARAMCO to do so, out of consideration.

The 'Onassis game', as presented here, serves very well to demonstrate the conditions under which a decision-maker might not be able to consider an objectively 'wise choice'. In this case the perceptual limitations imposed by Onassis' state of crisis, the extreme financial urgency of the question and the single minded pursuit of his aim (to save the agreement) caused an apparently 'rational' decision to be very detrimental to his own goals. If Onassis had diverted at least some of his efforts to the study of Saud's aims and choices he would have probably been able to foresee the dangers inherent in his refusal to renegotiate.

Our conclusions can also be expressed using the model of decision-making proposed by Radford [16]. He divides the process into three interrelated and cyclical phases:

(i) Information Gathering,

(ii) Strategic Analysis,

(iii) Interaction between Participants (Negotiation, Communication and Bargaining).

What this example demonstrates is the enormous influence that one party can exert on another in phase (iii) by affecting his analysis in phase (ii). Indeed, one can plausibly argue that often the most important function of negotiation, communication and bargaining is to shape the other parties' strategic models of the situation in such a way that their (voluntary) decisions accord with one's own aims. We can also see that phase (i) is shaped by (ii). The information one gathers depends on the model one starts off with; observation is always 'theory-directed'.

Ideally this case study should have shown whether the hypergame model gives accurate predictions of people's behaviour. But due to the historical nature of case studies, it could be objected that the games and stability criteria were chosen and constructed in such a way that they had to give the expected and desired results. Clearly it is impossible to completely disperse this suspicion, but it is hoped that the sometimes rather lengthy descriptions of players' perceptions do provide a basis for critical scrutiny of the problem formulation. No author can deny that his expectations, at least subconsciously, influence his findings. But by a clear statement of the assumptions the modelling process can be subjected to external critique. It was already mentioned above that the stability criteria were chosen mainly by virtue of their simplicity (not because they happened to give the desired results). This not only minimises the work for the research worker but it is probably quite a reasonable guide to human decision behaviour, even when the problem itself is complex. Lindsay and Norman, for example, state:

> "Several studies of chess experts... Suggest that chess masters use a number of heuristic rules to examine and select moves.... These are quite simple rules... of more importance, they succeed without requiring too much memory capability on the part of the player."
>
> [12, p. 556]

The criteria used here are such *simple rules*; but whether they can provide a more general description of decision behaviour remains to be seen.

In summary it can be said that in the context of this case study the hypergame proved to be a good model for decision situations involving misperceptions. A conventional game would perhaps been sufficient to show the finally achieved outcomes to be stable. However, it would not have been able to show any of the reasons behind their stability. The fact that, due to his information conditions, King Saud was able to anticipate Onassis' misperception gave his actions their justification. Only a model incorporating these misperceptions and the different information conditions is consequently able to give an adequate representation of the situation.

Undoubtedly the hypergame model runs into some of the problems of conventional games, such as the lack of a universally accepted solution concept and its 'static' character, but by using perceptual games as the basis of the model it also overcomes some of the major limitations of the traditional theory of games.

REFERENCES

1. BENNETT PG (1977 Toward a theory of hypergames. *Omega*, 5(6) 749–751.
2. BENNETT PG (1978) *Hypergame* Analysis. Department of Operational Research, University of Sussex, UK.

3. BENNETT PG & DANDO MR (1977) Fall Gelb and other games: a hypergame perspective of the fall of France, 1940. *Jl Confl. Res. Soc.* **1**(2), 1–33.
4. BENNETT PF & DANDO MR (1979) Complex strategic analysis. *Jl Opl Res. Soc.* **30**(1), 23–32.
5. BERRESFORD A & DANDO MR (1976) OR for strategic decision making. *Jl Opl Res. Soc.* **29**.
6. BRADY F (1978) *Onassis: an Extravagant Life*. Futura Publications, London, UK.
7. BRAMS SJ (1977) Deception in 2 × 2 games. *Jl Peace Sci.*
8. FRASER N, JACOBSON P, OTTAWAY M & CHESTER L (1977) *Aristotle Onassis*. Weidenfield and Nicholson, London, UK.
9. GIESEN MO (1978) The Hypergame Model—a Critical Examination Using Case-Studies. BSc. dissertation, Department of Operational Research, University of Sussex, UK.
10. HARSANYI JC (1968) Games with incomplete information played by 'Bayesian' players. *Mgmt Sci.* (Ser. A) **14** (i) 159, (ii) 320, (iii) 486.
11. HUXHAM CS, BENNETT PG & DANDO MR (1978) Building for scrap: a discussion of the shipping crisis and some analysis using hypergames. University of Sussex.
12. LINDSAY PH & NORMAN DA (1977) *Human Information Processing*. Academic Press, New York, USA (2nd Ed.).
13. LUCE RD & RAIFFA H (1957) *Games and Decisions*. John Wiley, New York, USA.
14. MIKDASHI Z (1966) *A Financial Analysis of Middle Eastern Oil Concessions*. FA Praeger, New York, USA.
15. NASH J. (1951) Non-cooperative games. *Ann. Math.* **54**, 286.
16. RADFORD KJ (1977) *Complex Decision Problems*. Reston Publishing Company.
17. SAYEGH KS (1968) *Oil and Arab Regional Development*. FA Praeger, New York.
18. VON NEUMANN J & MORGENSTERN O (1944) *Theory of Games and Economic Behaviour*. Princeton University Press, Princeton, USA. (2nd ed. Princeton 1953).

ADDRESS FOR CORRESPONDENCE: *Dr PG Bennett, University of Sussex, Department of Operational Reserach, Mantell Building, Falmer, Brighton BN1 9RF, UK.*

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 4, pp. 321–332
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/0801-0321$02.00/0

Inventory Management and Cost–Benefit Analysis: A Case Study[1]

STEVEN NAHMIAS

Stanford University

GEOFFREY LEVINE

University of Pittsburgh

(*Received August* 1978; *in revised form January* 1979)

This paper treats the problem of comparing various stocking and issuing policies for two radioactive isotopes, I-123 and I-131, which are used in nuclear medicine departments for the diagnosis of thyroid disease. The approach is based on using a combination of a cost–benefit analysis to compare the increased operating costs of stocking I-123 versus the increased disease risk of using I-131, and simulation to recreate the operation of a radio-pharmacy which stocks these drugs. The required inputs to the model include: (1) estimates of the disease risks associated with radiation absorbed doses from I-131; (2) estimates of the probability distributions of the appropriate latent periods; (3) empirical distributions of daily patient demand; (4) census data for estimating longevity and individual expected earnings and (5) cost data the stocking costs of the isotopes. The cost benefit results give issuing rules based on the age and sex of the patient, while the simulation is used to compare the effectiveness of the various ordering policies considered (the choice of a particular ordering policy may influence the risk of disease development).

I. INTRODUCTION

THE PURPOSE of this study is to present a specific inventory management problem which is faced by the nuclear medicine departments throughout the country, and demonstrate why the usual types of costs and profit and objectives which dominate inventory models are completely inappropriate in the context of this problem. Further we show how cost benefit analysis can be applied to the problem of comparing alternative stocking and issuing policies, which is, to our knowledge, the only inventory management model which utilizes this type of methodology.

The problem we consider involves the stocking of two radioactive isotopes, I-123 (Sodium Iodide 123) and I-131 (Sodium Iodide 131) which can be used interchangeably for the diagnosis of thyroid ailments including cancer. Although both of these isotopes may be used to satisfy the same demand, they differ markedly in their physical characteristics: I-123 has a physical half life of 13.3 hours (thus requiring daily shipments) while I-131 has a physical half life of 8.1 days; I-123 emits only gamma radiation while I-131 emits both gamma radiation and beta particles (the beta particles are the primary contributor to radiation absorbed dose); I-123 has a shelf life of only one day while I-131 has a shelf life of approximately 30 days; I-131 is obtained as a waste product of the fission reaction within a nuclear reactor and is thus relatively inexpensive to procure, while I-123 is produced by a particle accelerator process (and is not a waste product of

[1] Research partially supported by the National Foundation under Grant ENG 78-05928.

another reaction) and as a result is more expensive to procure.

The question we address is: what is the 'best' stocking policy for these two substitute products and what rules should be employed for issuing the two drugs. From the point of view of classical inventory control models, the answer is obvious. Since I-131 has a longer shelf life and is cheaper to procure than I-123, the minimum cost solution calls for exclusively using I-131. However the actual problem is far more subtle. The emission of the beta particles by the I-131 is itself a possible cause of considerable localized radiation exposure which may ultimately result in the development of thyroid ailments. Hence the use of the I-131 isotope for diagnostic purposes may have the effect of eventually inducing one of the variety of conditions it is intended to diagnose.

The exclusive use of I-123 would result in minimal disease risk since it emits (essentially) only gamma radiation. However, because of the uncertainties associated with patient demand, the extremely short life of I-123 and its high costs of procurement, the operating costs of using I-123 exclusively might be prohibitive for many hospitals. The problem is then to develop a quantitative model for balancing the additional costs of I-123 versus the reduced risk[2] of its possibly inducing a later thyroid ailment.

If we think of the decision maker for this problem as society as a whole rather than just the local hospital, then two types of costs to society will result if an individual should develop a thyroid ailment: the direct costs of medical care and the indirect costs of lost productivity. An optimal stocking policy for the two isotopes will balance off the additional costs of stocking I-123 with the expected future costs associated with the risks of using I-131.[3]

This problem appears to be extremely difficult to model mathematically. Although inventory models do exist for items which are subject to continuous exponential decay (see [6]) and fixed life perishability (see [19]), we are dealing with a two product pure substitute system in which both products are simultaneously decaying and possess a fixed shelf life. In addition, the objective function is based on balancing operating and disease risk costs. For these reasons a pure mathematical model appears difficult to develop. As an alternative, an approach based on monte-carlo simulation will be considered.

The following is a general outline of the methodology which is used in this study:

(1) Based on a number of previous studies reported in the literature, the risk probabilities associated with developing (a) thyroid cancer, (b) hypothyroidism and (c) a 'catch-all' group including benign goiters and nodules (herein referred to as nodularity for convenience) are estimated as a function of the age when received for a specified dosage in rads (units of radiation absorbed dose). The absorbed dose in rads is related to the units of radioactivity administered (in micro-curies) for a specified radioisotope. Additional information is used to determine the probability distribution of the latent period (i.e. the period of time that elapses from the time of irradiation to the time one of the conditions above is developed, or is diagnosed).

(2) Using these risk probabilities, the expected discounted cost of developing one of the conditions above is computed, based on the assumption of a 500 rad dose of I-131. Assuming that the risk probabilities are a linear function of the dosage, for a given age, these probabilities can be computed for other, more reasonable, lower dosages in the diagnostic range.

(3) When the expected loss due to I-131 exceeds the additional cost of I-123, then I-123 is more desirable. Included in this computation are costs of medical therapy and loss of income. A decision rule based on the age and sex of the patient is developed which specifies the optimal radioisotope to use.

[2] Commercially available I-123 is mildly contaminated with I-124, I-125, I-126, I-130 and I-131. However, we may ignore the risk associated with the contaminants as it is relatively small if the I-123 is administered on the day of calibration. Note that the disease risk due to pure gamma radiation was assumed to be zero for the purposes of this study.

[3] One must view this statement in the perspective of the problem itself. There is also a possibility that the disease will not be diagnosed or be diagnosed by other means such as surgery which has a higher risk of causing death or a later complication.

Vol. 7
1979

TABLE 1. THYROID CANCER CASES FOR CHILDREN UNDER 10 YEARS OF AGE

Study	Cases/10^6 persons/rad/year
Hempelmann [10] (Rochester series)	5.5
Hempelmann [11]	2.5
Latourette and Hodges [13]	2.2
Marshall Island Report [16]	1.8
Beach and Dolphin [1]	1.7
UNSCEAR [25]	0.9

(4) A simulation model is then developed to compare the performance of ten different ordering policies. Empirical data are used to generate both demand distributions and patient age. The patient priority class (inpatient or outpatient) is considered in the simulation.

II. DETERMINATION OF THE RISK PROBABILITIES

A basic assumption that is made in determining the probability that any of the disease conditions is developed by the patient is that the risk is a linear function of the absorbed radiation dosage (measured in rads). A similar assumption was also made in [2]. Since the risk associated with zero dosage is zero, the linearity assumption implies that only a single data point is required to determine the risk probabilities for any condition over the entire range of possible values of the absorbed radiation dosage.[4]

Two of the most significant studies which were used to determine the shape of the risk curve for thyroid cancer were those of Hanford *et al.* [9] and Socolow *et al.* [23]. Hanford followed the history of patients treated at a New York hospital for benign lesions during the period 1920–1950. Based on his data, the risk probabilities associated with a 1000 rad dose as a function of the age of the patient when the dose was received is given by the following:

Age	Risk (Probability of developing a benign lesion)
0–9	17%
10–19	10.8%
20–29	1.3%
30–39	1.7%
over 40	0%

Although Hanford found zero risk for patients over 40, Socolow's study of the survivors of the Hiroshima and Nagasaki blasts indicated that there was a positive risk for patients over 40.

A number of other studies were concerned with the incidence of thyroid cancer to children under 10 years old. Table 1 below gives the results of these studies.

The average of these observations is 2.43. Since this yearly data represents an approximate 20 year span, we multiply by 20 to obtain the overall risk per rad per 10^6 individuals, which will give a very conservative estimate of the true risk. For a dosage of 500 rads the risk per individual in the 0–10 year age category is then obtained as

$$2.43 \times \frac{1}{10^6} \times 20 \times 500 = 0.0243$$

A similar calculation based on the studies by others [1, 9, 12, 23–25] gives an average risk for *all* age groups as 1.4% for a 500 rad dose. Based on the observation that the risk curve is monotonically decreasing with age and the estimates of 2.43% as the average risk for the zero to ten age group and 1.40% as the average risk overall, we obtain an estimated risk curve of developing thyroid cancer for all ages at an absorbed dosage of 500 rads. The results are pictured in Fig. 1.

In order to determine the risk curve for nodularity, various studies [2, 3, 5, 21] were used and estimates of the average risks for the

[4] The linearity assumption has been traditionally used in radiation dose calculations, although this assumption as well as the assumption of no threshold (that is, zero risk for zero dose) have come under fire. We believe the model to be realistic for the purposes of this study.

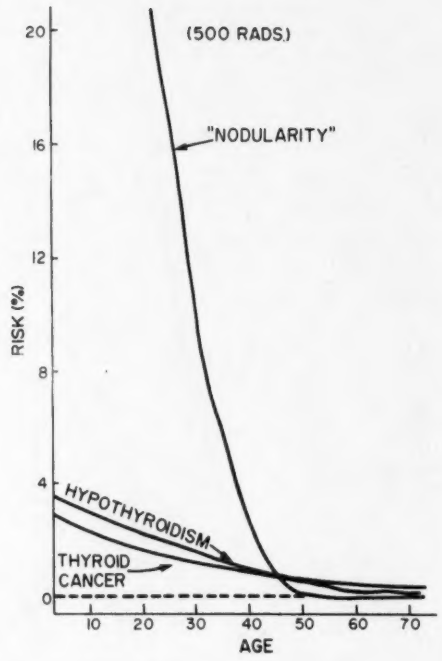


FIG. 1. *Disease risk for 500 rad dose.*

various age groups made in a fashion similar to that for thyroid cancer. In the case of hypothyroidism the relevant studies are those of the Marshall Island Report [16] and Green and Wilson [8].

The estimated risks for a 500 rad dosage for all three disease conditions are given in Fig. 1. Figure 1 should be interpreted as follows: if an individual of any age level specified on the abscissa receives an absorbed dosage of 500 rads then the probability that he contracts one of the three disease conditions specified during the remainder of his lifetime is given by the specified curves. Since risk is assumed to be a linear function of absorbed dosage, the risk associated with any dose from 0–500 rads may also be obtained from Fig. 1. For example, if an individual of age 40 receives a 25 rad dosage the likelihood of his developing a 'nodularity' condition during his remaining lifetime is approximately (0.03) × (25/500) or 0.0015.

It is interesting to note that our estimate of 2.43 cases per year per million persons receiving a dose of one rad is in remarkably close agreement to the later estimates reported at the 1976 Symposium on Radiation Associated Thyroid Disease [4]. In this report an estimate of three cases per year per million persons for each dosage of one rad was made for persons treated in childhood with x-rays for benign disorders (such as acne or enlargement of the thyroid or lymph nodes due to infection).

Similarly, we find close agreement with our estimates and those of the panel regarding the occurrence of benign nodules. (The report expresses risk in a different manner however: if 25 years ago the thyroid glands of 100 people were exposed to 300 rads each, then approximately nine will develop benign nodules on the thyroid.)

Our risk estimates also agree with the findings of the Bureau of Radiological Health (private communication to the second author).

III. ESTIMATION OF THE LATENT PERIOD PROBABILITY DENSITIES

The latent period is defined as the amount of time that elapses between the patient being irradiated and his contracting the particular disease condition considered. We will make two assumptions which appear to be fairly well supported by experience. One is that the latent period is independent of the magnitude of the absorbed dosage [10] and the other is that the latent period is independent of the age of the patient at the time of irradiation [9].

Since we treat the latent period as a continuous random variable, it is necessary to estimate the probability densities associated with the latent periods for the three disease conditions considered. Our estimates are extremely rough owing to the paucity of data available. In the case of thyroid cancer, Hempelmann's [10] study would suggest a mean period of about 20 years and a maximum of about 40 years. In the case of hypothyroidism, Green and Wilson [8] indicate that a major portion of the development occurs within one year while Silver [22] implies the latent period for hypothyroidism rarely exceeds four years. In the case of nodularity, Silver indicates that nodules are most likely to appear from two to five years after irradiation while the Marshall Island Report [16] indicates that nodularity is almost certain to appear within ten years.

Using this information, we have estimated the probability densities of the latent periods for each of the disease conditions considered

Vol. 7
1979

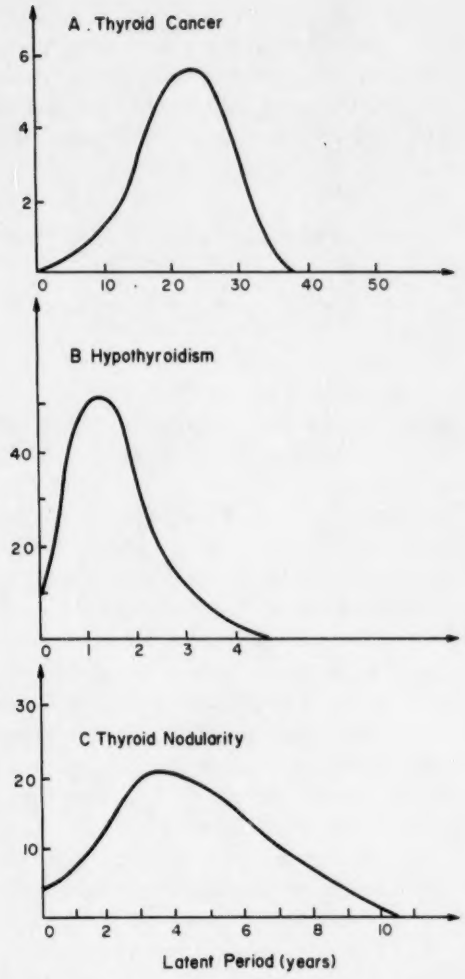


FIG. 2. *Estimate of probability densities of latent period for developing thyroid cancer, hypothyroidism, or thyroid nodularity following radiation exposure.*

which are pictured in Fig. 2. Each density is roughly symmetric, is unimodal and has a range of abscissa values chosen to be consistent with the above. Ordinate values are selected to guarantee that the area under each curve is one.

IV. DETERMINATION OF DISEASE COSTS

For each of the three disease conditions considered in this study there are both direct and indirect costs which must be taken into account. Direct costs are those required to actually treat the disease and include those costs incurred for surgery, medical tests, physicians' expenses, drugs, supplies, etc.

For thyroid cancer, one must include complete diagnostic radiology preliminary tests before surgery (including nuclear medicine and x-rays), thyroid gland surgery, and post surgery tests. The patient remains hospitalized for an average of twelve days and must return each year for a period of five years for follow-up tests. Based on data current in 1974, we obtain a direct cost of $2433.31 in the first year and $150.58 for each of the next five years. The present cost (relative to 1974) based on an inflation rate of 10% is $3050.72.

For nodularity, generally surgery or hospitalization is not required, but tests and drugs may be administered for a period of approximately five years from the time the condition is discovered. The estimated cost is $45.43 per year which results in a present direct cost for this condition of $231.70.

Hypothyroidism treatment includes a series of tests the first year and the administration of a thyroid hormone medicine for the remaining lifetime of the patient. Using the estimate of 35 years as average remaining lifetime (that is, the time from discovery of the disease until death) using census data reported in [17] one obtains from an estimated direct cost of $121.03 the first year and $2.50 for each following year a present direct cost (relative to 1974) of $153.42 for hypothyroidism.

Indirect costs include those costs to society which result from an individual developing one

TABLE 2. AVERAGE EARNINGS BY AGE* (1974)

Age	Male	Female
Under 14	0	0
14–19	1571	1324
20–24	4647	3291
25–29	8510	4162
30–34	10,217	4028
35–39	11,329	4177
40–44	11,814	4424
45–49	11,879	4621
50–54	11,298	4690
55–59	10,409	4634
60–64	9201	3929
65–69	6367	2807
70–74	4978	2514
75 and over	3922	2226

* The entries in this table were obtained from the 1970 Census Report [27] and extrapolated to 1974 assuming an increase of 3% per year.

of the three disease conditions considered. We will assume that the indirect costs are proportional to the loss of the individual's productivity which is measured by the loss in earnings for that individual. A common estimate of the constant of proportionality is 2.5. Average earnings for males and females as a function of age are estimated from the US census report [27], and given in Table 2. Hence the calculation of indirect costs depends upon the age and sex of the patient, at which point the condition is discovered as well as the specific condition itself.

Let x be the average daily earnings for any individual. Since thyroid cancer involves an initial therapy period of twelve days the first year, and one day per year for the next five years, one obtains an indirect cost of $\$30x$ the first year and $\$2.5x$ each following year. Note that we have applied the proportionality factor 2.5. Hence the total indirect cost for thyroid cancer is $\$40.25x$.

The therapy for nodularity involves a period of one day per year for a total of six years. The resulting indirect cost is $\$12.75x$. The therapy for hypothyroidism involves only one day of initial treatment. The resulting indirect cost is $2.5x$.

V. THE COST BENEFIT APPROACH

The proper manner in which to assess a dollar cost for risks associated with disease or loss of life is an issue which has never been truly resolved. For the purposes of this study we have chosen what is probably the most common method: the economic penalties to society for using I-131 (which is the same as the economic benefit of using I-123) is taken to be the sum of the expected costs of medical care (direct costs) and expected loss in future discounted earnings (indirect costs) which accrue as a result of the risks associated with the three disease conditions. Such an approach ignores the fact that there might be some cost levied for the suffering of the individual and the people around him and that people are generally risk averse in that they would be willing to pay a substantial premium to avoid such risks. Mishan [14, Chapter 8] cogently discusses this issue.[5] Because of the strong differences of opinion on this point, we feel that our approach is the most reasonable in the context of the problem being considered, and appears to be entirely consistent with the philosophy of the BEIR Report [2]:

> "The concept of balancing the risk of radiation exposure against the expected benefit has been well recognized and accepted, but no serious attempt has been made to evaluate both sides of the equation in any way that could lead to operational guidance."

This study represents perhaps the first such attempt. The use of cost–benefit analysis in health planning however is well known, see for example [7, 15, 20, 24].

The computation of the expected benefit due to I-123 is accomplished as follows. Future costs are discounted to the present assuming an interest rate of r. In addition, inflation is accounted for by allowing for both the average annual increase in the rate of medical costs, m (which at the time of this study in 1974 was about 4% per year), and the average annual increase in wages, s, which can be estimated from census data. The effective discount factor for medical costs is then $(1 + m)/(1 + r)$ and the effective discount factor for wages is $(1 + s)/(1 + r)$. The expected benefit from using I-123 (the expected cost from using I-131) is computed from the formula given in the appendix. Separate calculations are performed for men and women. This figure is then compared to the difference between the unit costs of I-123 and I-131 which for this study is assumed to be \$7.00, which includes the air freight daily shipping charge for I-123. The cost–benefit assumption is that I-123 is preferred if the expected benefit (given by the formula above) exceeds \$7.00 while I-131 is preferred if this benefit is less than \$7.00.

The expected loss for administering the inappropriate isotope is pictured for males in Fig. 3 for interests rates of $r = 0.03$, 0.07 and 0.11. Note that the loss is highest for the younger patient due to the increased risk of contracting one of the three disease conditions. The crossover points occur at ages 43, 34 and 27 for interest rates 0.03, 0.07 and 0.11 respectively. When the age is *less* than these crossover points, the expected benefit of I-123 exceeds the \$7.00 difference in unit costs and as a result I-123 is more desirable. However, if the

[5] We are grateful to an anonymous referee for bringing this reference to our attention.

Vol. 7
1979

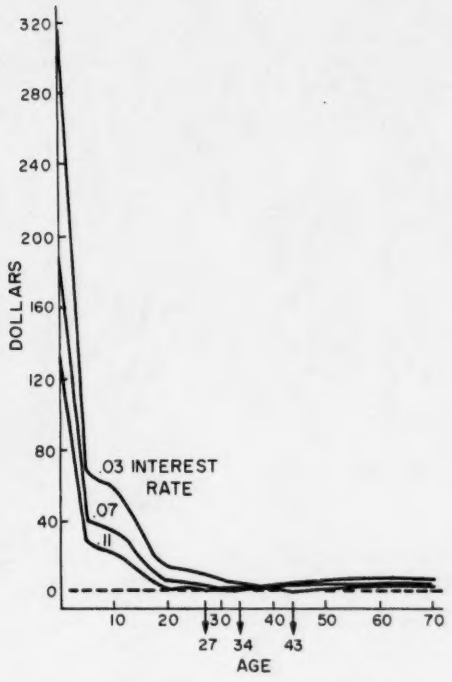


FIG. 3. *Expected loss for administering the alternate isotope.*

patient's age *exceeds* the crossover point, the expected benefit of I-123 is less than $7.00 and so I-131 is assumed to be more desirable. The graph pictures the absolute difference between the $7.00 additional procurement cost of I-123 and its expected benefit. A similar graph (not pictured) was developed for females. In that case the respective cut-off for choosing I-123 were 39, 28 and 23 for the respective interest rates of 0.03, 0.07 and 0.11.

A validation of our results is provided by the fact that these ages are consistent with the general practice in the Nuclear Medicine community. Generally I-131 is not given to individuals of child bearing age or younger when I-123 is available.

VI. SIMULATION ANALYSIS OF ORDERING POLICIES

The purpose of the simulation is to compare the performance of a number of different 'reasonable' ordering policies in terms of their costs to the patient and to the hospital. The probability distribution of daily demand is estimated from empirical demand data. These distributions are used in conjunction with the computer's random number routines to obtain sample realizations of daily demand. In a similar fashion, empirical data were also used to develop the probability distribution of the patient's age at time of diagnosis. (Sample data covered the period March–June 1974).

The ordering policies which were tested were of the following type: a standing order to 10, 15, 20 or 25 capsules per week of I-131 was considered in conjunction with daily orders of I-123 varying from 0 to 6. The total number of capsules of both types ordered each week was fixed at about 25 (to reflect the fact that weekly demand rarely exceeds this figure). The ten ordering policies considered are given in Table 3 below.

These particular policies were considered to be reasonable in view of the current types of operating characteristics of the system.

The simulation program also accounts for two patient priority levels: inpatient and outpatient. It is assumed that an inpatient demand will always pre-empt an outpatient demand since outpatients can be rescheduled with rela-

TABLE 3.

Policy	Standing order of I-131	Daily orders of I-123 M	T	W	TH	F
1	25/week	0	0	0	0	0
2	20/week	0	0	5	0	0
3	20/week	3	0	3	0	0
4	15/week	5	0	5	0	0
5	30/2 weeks	5	0	5	0	0
6	15/week	4	3	0	4	0
7	15/week	3	3	3	3	0
8	20/2 weeks	5	3	5	3	0
9	0	5	4	5	4	2
10	0	6	5	6	5	4

tively little additional cost to either the patient or the hospital.

Since I-123 has only a one day shelf-life and doses are administered in the morning, the level of radioactivity of an I-123 capsule is approximately the same each day it is administered. However, this is not the case with I-131. Since I-131 is continuously decaying over its 31 day shelf life, its radioactivity level at time t, say A_t, is given by the standard decay equation $A_t = A_0 \exp(-\lambda t)$ where A_0 is the initial radioactivity and $\lambda = 0.00036$ is the decay rate constant (assume t is measured in hours). Depending on the age of I-131 when administered, more than one capsule may be needed to obtain the required dosage in microcuries. The object is to administer a dosage of sufficient radioactivity to obtain a good quality study, while minimizing the radiation absorbed dose (rads) to the patient, particularly to young patients. A poor image requires readministering a second larger dose.

In order to determine the correct value of the absorbed dosage in any particular case, it is necessary to obtain the correct relationship between the number of rads of radiation absorbed and the radiation level of the isotope (which is measured in microcuries (μCi) and these are the units used throughout in this paper) as a function of the age of the patient. This relationship is given in Table 4 below. In addition, it is necessary to know the relationship between the age of the patient and the dosage generally given. This relationship which is pictured in Fig. 4 was observed to vary roughly linearly from a dose of approximately 35 at age zero to a maximum of 100 at age 20 and beyond.[6]

To demonstrate how these relationships are used, we consider a simple example. Suppose a shipment of I-131 capsules have been in storage for 22 days and each capsule is initially calibrated at 100. Since these capsules have been in storage for $t = 24 \times 22 = 528$ hours, the radioactivity level of these capsules is $I_{528} = (100) \cdot \exp(-0.0036 \times 528) = 15$. Suppose the patient is 13 years old. Then the estimated level of radiation required for the study would be 77 $(35 + 100\text{-}35/20 \times 13)$ (see Fig. 4). This would call for administering five I-131 capsules that are 22 days old. From Table 4 this results in an absorbed radiation dose of 196 rads.[7]

TABLE 4.

Age of patient	No. rads/μCi absorbed by thyroid
0	32.00
1	10.00
2	8.58
3	7.15
4	5.73
5	4.30
6	4.08
7	3.86
8	3.64
9	3.42
10	3.20
11	2.98
12	2.76
13	2.54
14	2.32
15	2.16
16	1.88
17	1.66
18	1.44
19	1.22
20 and over	1.00

ASSUMPTIONS MADE IN THE SIMULATION

When both I-123 and I-131 were available, the results of the cost–benefit analysis were used to determine which isotopes to administer. However, if only one of the two were available the following rescheduling criteria were used:

(1) If I-131 was called for and only I-123 was available, then I-123 was used.

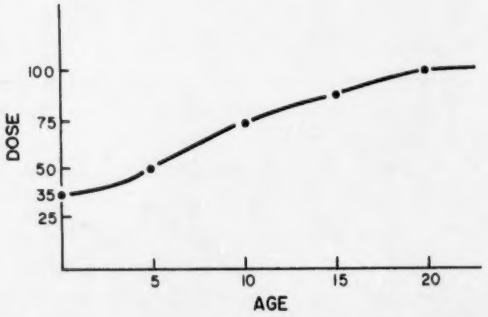


FIG. 4. *Dose per patient (I-131 μCi per patient) (estimated 1974 clinical practice) versus age at administration.*

[6] The choice of administered doses assumes a patient would receive a thyroid scan (an imaging procedure) rather than just a thyroid uptake study which would only require about one quarter of the scan dosage. Also we have not considered the possibility of multiple or follow-up doses.

[7] This example is meant for illustrative purposes only. It is relatively rare that as many as five capsules would ever be actually administered.

Vol. 7
1979

(2) If I-123 was called for and only I-131 was available then:

(a) If the patient was a priority 1 patient the I-131 was used.
(b) If the patient was a priority 2 patient, then he was rescheduled only if his expected loss exceeded a pre-specified tolerance. Tolerances of $10, $20 and $30 were considered.

(3) If neither I-131 nor I-123 were available, then the patient was rescheduled regardless of his priority status. That is, excess demand is assumed to be fully backordered. This is entirely consistent with current practice which does not allow for lateral resupply of isotopes among hospitals.

Additional inputs to the simulator were the probability distributions of daily demand for each isotope. These were estimated by analysis of patient records at the Presbyterian University Hospital, Pittsburgh, PA, and histograms for the number of units demanded as a function of the day of the week were generated. The age distribution of patients was also generated from patient records.

The process was simulated for 500 weeks of operation. Statistical tests indicated this was a sufficiently long running period to eliminate the transient effects of initial conditions while providing reliable estimates of the outputs of the simulation. For each fixed ordering policy and interest rate, the outputs of the simulator included: (1) the total dollar loss to the patient, (2) the total dollar cost to the hospital, (3) the percentage of each isotope which perished (excluding loss to decay) and (4) service levels for each priority patient (that is, what percentage of each type of patient could not be satisfied by demand).

The output of the simulation program assuming an interest rate of 7% and a rescheduling tolerance of $10.00 (as described above) is given in Table 5. Note that policy 1 which calls for exclusive use of I-131 with a standing order of 25 capsules per week resulted in minimum cost to the hospital of $8228.80 but a very high patient loss of $57,772.00 while policies 9 and 10 which call for using only I-123 resulted in higher costs to the hospital and relatively lower losses to the patient. The policy which resulted in lowest overall dollar losses was policy 7 which calls for a daily standing order of 3 capsules of I-123 (excepting Friday) and a weekly order of 15 I-131 capsules. It is interesting to note that this policy resulted in lower patient losses as well as lower hospital costs over policies 9 and 10.

Note that in Table 5, the percentage waste is broken down into two categories: 'waste' due to the use of a standard capsule which is required during the uptake and scan procedure in addition to the administered capsule, and the losses due to perishability.

The next two columns labelled '% given to patient' represent the percentages of each isotope which are actually used to satisfy demand. The next two columns give a measure of the service which results from each of the policies. For example, for policy 1, 18% of the inpatients and 30.08% of the outpatients had to be rescheduled due to the fact that no drug was available when it was demanded. Note that overall policy 7 seemed to give the best service performance as well as minimum overall loss. The final column represents the cost absorbed by the hospital when using each of the ordering policies. In general the hospital costs increased with the proportion of the quantity of I-123 ordered.

VII. DISCUSSION OF RESULTS

All of the numbers obtained here must be interpreted with care and the results must be viewed in the context of the political and operational constraints which are present in the hospital environment. In particular, the costs associated with any inventory policy are quite sensitive to values chosen for the costs of procurement and transportation. Many costs are difficult to account for in the model. One is the additional costs that arise from difficulties of transporting radioactive materials. Another is the costs of patient rescheduling. For this reason the quality of service plays an important role in policy comparisons.

Another important issue is the cost–benefit assumption that one accepts that radioisotope which results in minimum expected loss to oneself even if that choice involves choosing I-131 which may have a disease risk associated with it. However, in terms of disease risk, it is more likely that people would be 'risk averse' and

TABLE 5. SAMPLE OUTPUT OF SIMULATOR[7]

Policy	Cost per capsule to hospital		Total dollar loss	Dollar loss for		I-123			I-131			% Issued to patients		% In-patients rescheduled	% Out-patients rescheduled	$ Cost to hospital
	I-123	I-131		Patient	Hospital	Waste (%)	Standard (%)	Perish (%)	Waste (%)	Standard (%)	Perish (%)	I-123	I-131			
1	—	1.15	66000.3	57772.0	8228.8	—	—	—	57.2	4.0	53.2	0	100	0.18	30.03	6357.5
2	5.75	1.25	43498.5	29508.3	13990.3	43.1	20.0	23.1	62.4	5.0	57.4	27.4	72.6	0.14	21.55	9136.5
3	7.08	1.25	37491.5	21107.3	16384.2	38.9	33.3	5.5	65.0	5.0	60.0	34.4	65.6	0	13.62	12811.0
4	5.75	1.42	40048.2	20617.4	19430.9	42.8	20.0	22.8	67.0	6.7	60.3	53.6	46.4	0.08	13.87	11362.5
5	5.75	1.08	51842.1	26144.4	25697.7	42.2	20.0	22.2	83.5	1.7	81.4	53.9	46.1	0.08	13.13	16717.0
6	6.48	1.42	41060.1	17593.4	23466.7	45.3	27.3	18.0	69.1	6.7	42.4	56.5	43.5	0	8.41	17185.0
7	7.08	1.42	34943.3	7699.4	27243.9	45.7	33.3	12.4	73.7	6.7	67.0	62.3	33.7	0.03	3.77	22462.5
8	6.25	1.25	44333.5	9908.5	34425.0	46.5	25.0	21.5	89.3	2.5	86.8	80	20	0.08	4.39	24532.0
9	6.25	—	39340.2	10552.7	28787.5	46.1	25.0	21.1	—	—	—	100	0	7.94	23.8	16651.0
10	5.67	—	54827.2	10951.6	43875.6	59.5	19.2	40.3	—	—	—	100	0	0.65	0	28989.0

Running time = 500 weeks. Interest rate = 7%. Rescheduling tolerance = $10.00 (see text for explanation).

might prefer to pay additional money to offset a small but positive risk. In addition, the actual cost–benefit calculations are based on our estimates of disease risk and the probability distributions associated with the latent periods. Our estimates are extremely rough owing to the very skimpy data available. We would caution the reader to return to the original references before attaching any undue significance to the estimated risk probabilities given in Fig. 1 or the latent period distributions in Fig. 2.

The primary contribution of this study is that it outlines a methodology for dealing with the difficult trade-off of disease risks versus operating costs which occurs so often in health systems management. This study outlines a concrete method of computing a dollar loss associated with those risks. It also shows what we believe is the first example of cost/benefit analysis applied to the problems of issuing and ordering in an inventory system.

As a result of this study, the second author, who is responsible for ordering I-123 and I-131 for a centralized radiopharmacy serving many hospitals in the Pittsburgh area, has significantly altered the ordering policy for these two isotopes, resulting in a proportionately increased usage of I-123 after it became commercially available.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contributions made by Yoon Choy, Ranjit Gangopadyhay, Lalitkumar Shah and Richard Stephen.

REFERENCES

1. Beach SA & Dolphin GW (1962) A study of the relationship of X-ray dose delivered to the thyroids of children and subsequent development of malignant tumors. *Phys. Med. Biol.* **6**, 583–598.
2. BEIR (Biological Effects of Ionizing Radiation) (1972) The effects on populations of exposure to low levels of ionizing radiation. Report of Advisory Committee, *Nat. Acad. Scis, Nat. Res. Coun.*, Washington, DC. November.
3. Conard RA, Dobyns BM & Sutrow WW (1970) Thyroid neoplasia as late effects of exposure to radioactive iodine in fallout. *Jl Am. Med. Assoc.* **214**, 316–324.
4. Chapman EM (1977) Final panel report V; external radiation and thyroid cancer incidence, in *Radiation Associated Thyroid Carcinoma* (Ed. Degroot LJ), 507–508. Grune and Stratton, New York, U.S.A.
5. Dolphin GW (1968) The risk of thyroid cancers following irradiation. *Hlth Phys.*, **15**, 263–274.
6. Emmons H. (1968) A replenishment model for radioactive nuclide generators. *Mgmt Sci.* **14**, 263–274.
7. Fuchs T (1972) Essays in the Economics of Health and Medical Case. *Nat. Bur. Econ. Res.* New York, USA.
8. Green & Wilson GM (1964) Thyrotoxicosis treated by surgery or iodine-131 with special reference to development of hypothyroidism. *Brit. Med. Jl*, **1**, 1005–1010.
9. Hanford JM, Quimby EH & Frantz VK (1962) Cancer arising many years after radiation therapy: incidence after radiation of benign lesions in the neck. *Jl Amer. Med. Assoc.* **181**, 132.
10. Hempelmann LH, Pifer JW, Burke GJ, Terry R & Ames WR (1967) Neoplasms in persons treated with X-rays in infancy for thymic enlargement. Report of Third Follow-up Study. *Jl Nat. Cancer Inst.* **38**, 317–341.
11. Hempelmann LH (1968) Risk of thyroid neoplasms after irradiation in childhood. *Science*, **160**, 159–163.
12. ICRP publication 8 (1966) *The Evaluation of Risks from Radiation.* Pergamon Press, Oxford, UK.
13. Latourett HB & Hodges FJ (1959) Incidence of neoplasm after irradiation of thyroid region. *Amer. Jl Roentgenology Rad. Ther.* **82**, 667–677.
14. Layard R (1972) *Cost Benefit Analysis*, Selected Readings. Penguin, Middlesex, UK.
15. Lesourd DA, Fogel ME & Johnston DR (1968) Benefit–cost analysis of kidney disease programs, *US Dept Hlth, Educ. & Welfare Res. Triangle Inst.*
16. Marshall Island Report (1971) Thyroid disease as a late sequel of radioactive fallout. Flint Labs, Morton Grove, Illinois, USA.
17. Metropolitan Life Ins. Co. (1973) Statistical Bulletin.
18. Myers WG (1973) I-123 for applications in diagnosis. *Radiopharmaceutical and Labelled Compounds*, **1**, 249–254.
19. Nahmias S (1975) Optimal ordering policies for perishable inventory—II. *Ops Res.* **23**, 735–749.
20. Rice DP (1966) Estimating the cost of illness. *Hlth Econ. Ser.* **6**, PHS Pub. No. 947–8, Washington, DC, USA.
21. Sheline GE, Lindsay S, McCormack KR & Gallante M (1962) Thyroid nodules occurring late after treatment of thyrotoxicosis with radioiodine. *Jl Clin. Endocr.-Metab.* **22**, 8–18.
22. Silver S (1962) Radioactive Isotopes in Medicine and Biology, 2nd edn, 127–132, 102–105. Lea and Febiger, Philadelphia, USA.
23. Socolow EL, Engel E, Mantooth L & Stanburg JB (1964) Chromosomes in human thyroid tumors. *Cytogenetics* **3**, 394–415.
24. Welsbrod BA (1961) *Economics of Public Health*, 30–32. University of Pennsylvania Press, Philadelphia, USA.
25. Wood JW, Tamagaki H, Sato S, Sheldon W, Archer PG, Hamilton HB & Johnson KG (1969) Thyroid carcinoma in atomic bomb survivors. *Amer. Jl Epidemiol.* **89**, 4–14.
26. UNSCEAR (Untd Nats Sci. Comm. on the Effects of Atom. Rad.) (1964) Report Supplement **14**, New York, USA.
27. US Census Report (1970).

ADDRESS FOR CORRESPONDENCE: *Professor Steven Nahmias, Department of Operations Research, Stanford University, Stanford, CA* 94305, *USA.*

APPENDIX

The total expected benefit of I-123 for a patient x years old (which is equivalent to the total expected loss of

I-131 to a patient x years old) is given by the formula:

$$\sum_{i=1}^{3} P_i(x) \sum_{j=1}^{T_x} h_{ij} \cdot \left\{ c_{ij} \cdot \left(\frac{1+m}{1+r} \right)^j + d_{ij} \left(\frac{1+s}{1+r} \right)^j \right\}$$

where

$P_i(x)$ = risk probability of disease condition i for a fixed absorbed dosage in rads received by a patient x years old (see Fig. 1)

T_x = expected remaining lifetime of a patient x years old

h_{ij} = probability that patient develops disease i after j years; that is the latent period probability distribution for disease i, (see Fig. 2)

c_{ij} = present direct cost of disease condition i to a patient $(x + j)$ years old

d_{ij} = present indirect cost of disease condition i to a patient $(x + j)$ years old

m = average annual increase in medical costs

s = average annual increase in wages

r = interest rate

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 4, pp. 333 to 338
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0801-0333$02.00/0

Analysis of the Choice Between Domestic and Foreign Loan Facilities

NIGEL MEADE

Imperial College, University of London

MALCOLM J FINNEY

JF Chown & Co Ltd, London

(*Received December 1978; in revised form February 1979*)

When a corporation borrows from a non-domestic source, such as the Eurobond market, it is obliged to make interest payments in a foreign currency at pre-determined periods throughout the life of the loan. By doing this the corporation incurs an exchange risk which may outweigh any advantages of borrowing foreign currencies such as lower interest rates. This note describes a criterion which assists the corporate treasurer in his choice of currency by showing the exchange rates at the end of the life of the loan implied by each choice.

INTRODUCTION

WHEN A corporate treasurer wishes to borrow over a long term, he has a choice of a number of currencies with associated interest rates. As a general rule, the more stable currencies offer lower interest rates than the less stable. If the corporation's domestic currency is one of the less stable ones (such as £), the treasurer's choice is likely to be between a high interest loan in his own currency, or a lower interest loan in a more stable foreign currency (such as Swiss Francs). The choice of the lower interest foreign loan implies that the exchange rate between the relevant currencies is not expected to change sufficiently for the effective payments of interest and principal to exceed those of the domestic loan. The choice can be formalised by computing a "breakeven" exchange rate at which the foreign and domestic loans would be equivalent.

It should be emphasised that the problem considered is that of making a choice between two or more fixed interest loans over specific terms, the Eurobond is an example of the type of loan instrument involved.

The use of the breakeven rate as a decision criterion for the corporate borrower was criticised by Finney and Meade [2] and an improved criterion was outlined. The purpose of this note is to describe the improved criterion in detail.

BREAKEVEN DECISION CRITERION

This criterion is nothing more than a straightforward internal rate of return calculation, it was described by Chown [1] in 1973.

Consider first a domestic loan over N years with an interest coupon of 100 $c_A\%$ *pa* with the interest payable n times *pa*, the cost of this loan will be the solution x to this equation:

$$1 - \sum_{i=1}^{Nn} \frac{(c_A/n)}{(1+x)^i} - \frac{1}{(1+x)^{Nn}} = 0 \qquad (1)$$

Obviously in this case the cost of the loan x is (c_A/n) domestic money units (dmu) paid n times yearly.

Secondly consider a loan in a foreign currency over the same period with the same timing of interest payments but with a coupon of 100 c_B% *pa.* When this loan is taken out the exchange rate is E_0 foreign money units (fmu)/dmu. It is also assumed that the relative depreciation rate between the currencies is r_b every $1/n$ of a year. The cost of the foreign loan is the solution y to this equation:

$$1 - \sum_{i=1}^{Nn} \frac{c_B/n}{(1 + r_b)^i} \cdot \frac{1}{(1 + y)^i} - \frac{1}{(1 + r_b)^{Nn}} \cdot \frac{1}{(1 + y)^{Nn}} = 0 \tag{2}$$

For the costs of the foreign and domestic loans to be equal:

$$y = x = c_A/n$$

Thus substituting in (2):

$$1 - \sum_{i=1}^{Nn} \frac{c_B/n}{(1 + r_b)^i (1 + c_A/n)^i} - \frac{1}{(1 + r_b)^{Nn}(1 + c_A/n)^{Nn}} = 0 \tag{3}$$

Thus for the cost of the loans to be equal the depreciation rate r_b must satisfy (3), that is at the end of N years the exchange rate must be $E_0(1 + r_b)^{Nn}$ fmu/dmu. The solution to (3) is simpler than it at first appears,

$$r_b = \frac{c_B - c_A}{n + c_A},$$

if r_b takes this value, the cost of the foreign loan is equal to that of the domestic loan and the end of term exchange rate $E_0(1 + r_b)^{Nn}$ is called the breakeven rate.

The treasurer bases his choice of loan on his view of the likely exchange rate at the end of the term of the loan in relation to the breakeven rate. Thus the breakeven rate is a simple and easily computed decision criterion.

However the assumption of a uniform geometric depreciation of relative currency values is implicit in this calculation. A study of any exchange rate between two currencies soon demonstrates that the rates fluctuate unevenly rather than follow a smooth path. In Fig. 1 the exchange rate of £ against the Deutschmark (DM) is plotted at 6 month intervals and the uniform depreciation rate over the same period is superimposed. The question that presents itself is to what extent do irregular changes in exchange rates undermine the usefulness of the breakeven rate.

SENSITIVITY OF BREAKEVEN EXCHANGE RATES

To measure the effect of the irregular fluctuations in exchange rates, a set of random

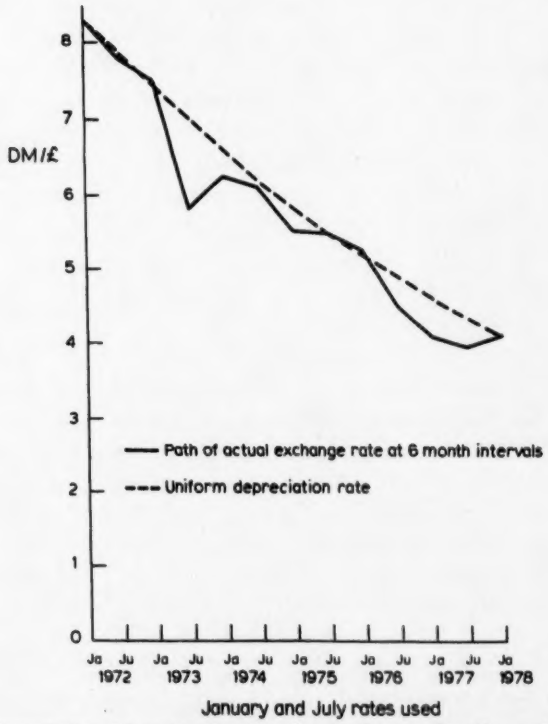


FIG. 1. *Depreciation of £ against the Deutschmark. (Last working day exchange rates)*

Vol. 7
1979

TABLE 1.

Period of loan (years)	1	2	3	4	5	10	15
Interest rates % p.a. (payable twice yearly)							
£	8.50	8.75	9.50	10.00	10.25	11.00	15.00
DM	3.25	4.00	4.75	5.25	5.50	5.50	5.75

paths to the breakeven rate are generated and the effective costs of the foreign loans are calculated. DM loans (foreign) were compared with £ loans (domestic) over the terms and using the rates shown in Table 1.

In order to generate the random paths from the initial rate to the breakeven rate, a mathematical model of the exchange rate movements must be hypothesised. No such models were found in the literature, however Granger and Morgenstern [3] demonstrated that a random walk model was appropriate for security prices on the New York Stock Exchange. As both security prices and exchange rates are decided by similar market forces the suitability of this model was tested for a set of seven major exchange rates. The results of this test provided no evidence to dispute the hypothesis of a random walk (copies of these results can be obtained from Meade). The random walk model is

$$\log E_i = \log E_{i-1} + \epsilon_i$$

where

$$E(\epsilon_i) = 0$$

and

$$\text{cov}(\epsilon_i, \epsilon_{i-s}) = 0$$

for

$$s \neq 0,$$

and E_i is the exchange rate at time i. The period to period depreciation is r_i where

$$1 + r_i = E_i/E_{i-1}$$

Since the exchange rate at the end of the period is fixed at the breakeven rate, this determines the mean value for r_i in the simulation of the paths of the exchange rate. In the simulation Normality of ϵ_i is also assumed, this can be justified by the Central Limit theorem in that the change in the exchange rate considered is a sum of many smaller fluctuations. Thus the intermediate exchange rates were generated using

$$\log(1 + r_i) \sim N(\mu, \sigma^2)$$

where

$$\mu = \log(1 + r_b)$$

and σ^2 is estimated from data. Using the Bank of England data illustrated in Fig. 1, the value of σ^2 for depreciation of £ against DM over 6 month intervals (to coincide with the loan requirements) from January 1972 to July 1978 was found. This value, $\sigma = 0.085$, was used to simulate exchange rate changes over the term of the loan, starting at the same initial rate and finishing at the appropriate breakeven rate. For each simulation run the effective interest cost was calculated, that is, the value of y in equation (2). The mean effective costs for each term of loan and 80% confidence intervals are shown in Fig. 2. One would expect the mean effective cost to be close to that of the domestic loan since each realisation of the cost of the foreign loan assumes the final breakeven rate is realised. This is the case, but the variation in the effective cost of the foreign loan is significant and should not be ignored. Thus

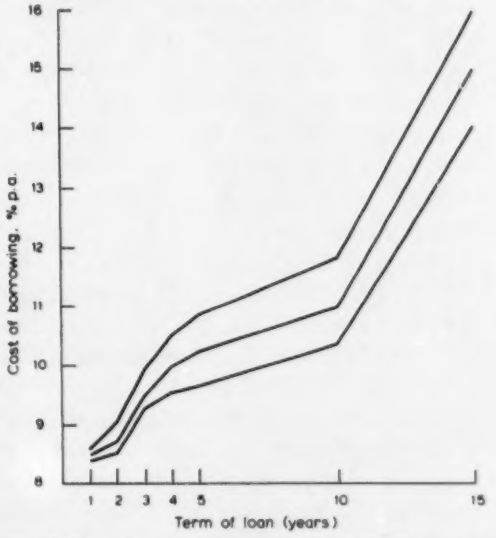


FIG. 2. *Domestic cost of borrowing* (£) *and 80% confidence intervals for effective cost of a D.Mark loan given end of term breakeven rate.*

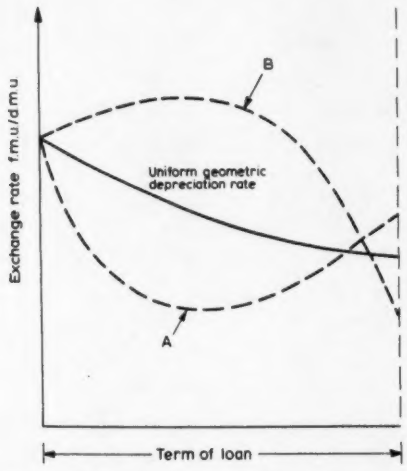


FIG. 3. *Uniform breakeven rate and two depreciation rates leading to equivalent costs of borrowing.*

the case for the use of the breakeven rate as a sole decision criterion is seriously weakened by the variability of exchange rate changes and it is clear that consideration of the variation of cost should be included in a modified decision criterion.

MODIFICATION OF THE DECISION CRITERION

The variations in the depreciation rate mean that the breakeven exchange rate will also vary. That is, the cost of the foreign loan will be equal to that of the domestic loan with a final exchange rate other than the uniform breakeven exchange rate. Thus in order to improve the decision criterion it should be possible to compute the probability distribution of the breakeven exchange rate using the information about the variations in the exchange rate.

The diagram in Fig. 3 illustrates the way in which different paths can lead to different breakeven exchange rates. Using uniform breakeven rate as the standard for comparison; for curve *A*, the currency depreciates faster than the uniform rate, thus interest rates are initially cheaper and for the effective cost of the loan to equal that of the domestic loan the resulting payment of the principal must be higher than that of a loan following the uniform depreciation rate. For curve *B* the reverse is the case, comparatively high interest payments must be offset by a lower repayment of principal.

The breakeven exchange rates further from the uniform breakeven rate are the results of more extreme paths, that is those that depart most from the uniform depreciation. The purpose of the rest of this section is to describe a convenient form of extreme path which can be used to generate the probability distribution of breakeven exchange rates.

To define the extreme outcome, an upper or lower confidence interval for the change in exchange rate over time is used, bearing in mind that the starting point E_0 and the finishing point E_{Nn} are both defined. Continuing the assumption that the change in the exchange rate is lognormally distributed, the upper confidence limit on the cumulative exchange rate at time t is as follows:

$$Pr\left(\sum_{i=1}^{t} \log(1 + r_i) > t\mu + k\sqrt{t}\sigma\right) = 1 - \Phi(k) \quad \text{for } t > 0$$

As the exchange rate changes are constrained to terminate at E_{Nn}, another upper confidence limit on the cumulative exchange rate change is:

$$Pr\left(\sum_{i=1}^{t} \log(1 + r_i) > t\mu + k\sqrt{(Nn - t)}\sigma\right) = 1 - \Phi(k) \quad \text{for } t < Nn$$

Similar expressions exist for the lower confidence limits.

The extreme path chosen which results in equality of cost of foreign and domestic loans is defined by these confidence limits. The final exchange rate is E_{Nn} and the intermediate rates are determined by

$$E_i = E_0 \exp\left[i \log(1 + r) + k\sigma\sqrt{\min(i, Nn - i)}\right] \qquad (4)$$

where

$$(1 + r)^{Nn} E_0 = E_{Nn}$$

Upper and lower extreme paths like curves *A* and *B* in figure 3 are defined by taking positive and negative values of k.

Equation (4) defines the shape of the path but the final rate E_{Nn} is the solution to this equation:

$$-1 + \sum_{i=1}^{Nn} \frac{c_B/n}{(1 + c_A/n)^i (E_i/E_0)} + \frac{1}{(1 + c_A/n)^{Nn} (E_{Nn}/E_0)} = 0 \qquad (5)$$

where E_i is defined in (4). Equation (5) is similar to equation (3) but the uniform depreciation rate is replaced by the extreme path depreciation rate.

For a given pair of currencies the value of k determines the probability that a sequence of exchange rates starting at E_0 and finishing

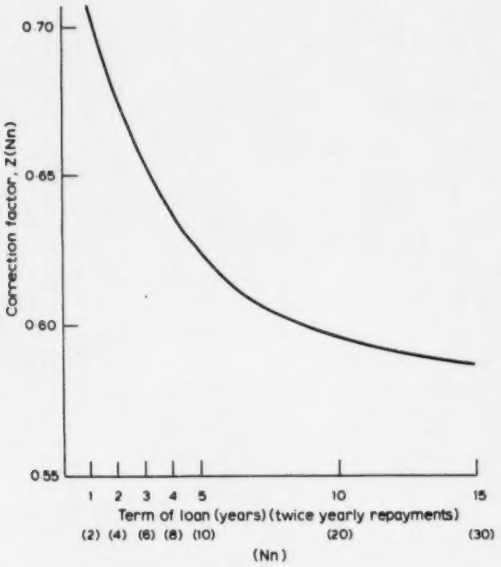


FIG. 4. *The function of* $Z(Nn)$

at E_{Nn} will result in a foreign loan costing more than a domestic one. That is, if F_i is a realisation of the exchange rate at time i, constrained only by $F_0 = E_0$ and $F_{Nn} = E_{Nn}$, then the effective cost of a foreign loan is C where C is the solution of (6):

$$-1 + \sum_{i=1}^{Nn} \frac{c_B/n}{(1 + C/n)^i (F_i/E_0)} + \frac{1}{(1 + C/n)^{Nn} (E_{Nn}/E_0)} = 0 \tag{6}$$

The probability of the event $C > c_A$ is a function of k and the conditions of the loan c_B and Nn,

$$\text{i.e. Prob } (C > c_A) = Q(k, c_B, Nn)$$

where Q is a function whose structure is to be determined. Values of C were found for randomly generated paths for F_i, for different values of Nn and c_B. No significant effect on $Q(.)$ was found due to c_B, but the number of payments of interest, Nn, did have an effect. The function $Q(k, Nn)$ was found to be adequately described in terms of the Normal distribution and an empirical function $Z(Nn)$ as follows

$$Q(k, Nn) \simeq 1 - \Phi\left(\frac{k}{Z(Nn)}\right). \tag{7}$$

The form of $Z(Nn)$ was found by experimentation and is shown in Fig. 4.

Thus an upper and lower confidence limit on the end of term breakeven exchange rates can be found using the value of σ and specifying $\alpha = Q(k, Nn)$, where 2α is the probability of a breakeven occurring outside the limits. For a given value of α, the value of k is determined from (7) and the values of the upper and lower limits form (4) and (5).

EXAMPLE AND CONCLUSIONS

The recommendation resulting from this analysis is that the single point breakeven exchange rate is replaced by a band of exchange rates where breakeven is possible. The point is best made by the example shown in Fig. 5 for a treasurer whose domestic currency is Sterling considering a Deutschmark borrowing using the rates shown in Table 1. If the end of term rate falls in region *A*, then the DM loan will be cheaper with a probability of at least 90% ($\alpha = 10\%$) in region *C* the £ loan will be cheaper with a probability of at least 90%. In region *B* above 'the' breakeven rate the DM loan will be cheaper with a probability between 50% and 90%; below it £ loans will be cheaper with probability between 50% and 90%.

The same procedure can be carried out for any value of α, so that the probability distribution of the breakeven rates can be defined as precisely as required.

By using this modified criterion the treasurer will have a clearer representation of the effects associated with the end of term exchange rates than the somewhat deceptive picture given by the single breakeven rate.

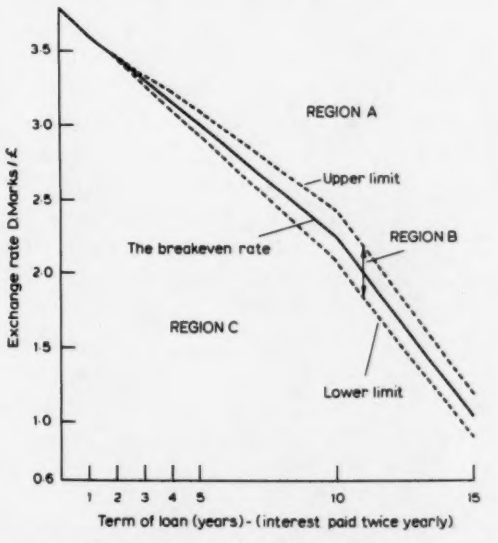


FIG. 5. 80% *limits on breakeven rates for D.Mark versus £.*

REFERENCES

1. CHOWN JF (1973) Financial Times Tax Newsletter (April).
2. FINNEY M & MEADE N (1978) A practical approach to corporate borrowing and exchange risk. *Euromoney* (October), 191–197.
3. GRANGER CWJ & MORGENSTERN O (1970) *Predictability of Stock Market Prices*. Heath Lexington, Massachusetts, USA.

ADDRESS FOR CORRESPONDENCE: *Dr. Nigel Meade, Department of Management Science, Imperial College of Science & Technology, Exhibition Road, London SW7 2BX, UK.*

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 4, pp. 339 to 344
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0801-0339$02.00/0

A Portfolio Model for Foreign Exchange Exposure Management

LUC A SOENEN

Eindhoven University of Technology, The Netherlands

(*Received November 1978; in revised form January 1979*)

This paper summarizes the results of our research into applications of decision analysis and portfolio theory to the management of foreign exchange exposure. In contrast with much current practice in foreign exchange management, the portfolio approach takes into explicit consideration the inherent relationships among the currencies in the company's foreign currency portfolio. The hedging model developed in this article traces out an 'efficient frontier' or trade-off curve between expected value and variance of the foreign currency portfolio at the end of the planning period. In doing so, the model chooses the optimal amount and method of hedging for each currency in the portfolio.

INTRODUCTION

CURRENT exchange exposure management is usually executed on a currency by currency basis failing to recognize the statistical relationships among the changes in the exchange rates of the currencies maintained by the company, not to mention exploiting the reduction of exchange risk holding a diversified portfolio of currencies. Although traditional exposure analysis has been conducted for individual currencies separately, portfolio theory has taught us that individual risks should not be measured independently. Because of diversification, the risk of a portfolio composed of various currencies is not equal to the sum of the risks of each currency. Just as unsystematic risk can be reduced via diversification by the number of stocks in an investment portfolio, so exchange risk can be reduced in a diversified portfolio of currencies. Consideration of the statistical relationship among currencies allows exposure netting as a technique to reduce exchange risk. If one believes that exchange rate movements of two currencies are highly positively correlated, then a short position in one currency will largely offset a long position in the other currency. If the currencies are negatively correlated, long positions in both currencies (or short positions) will tend to balance each other out.

Traditional techniques of foreign exchange risk management (monetary balance, swaps, forward markets etc.) do not explicitly recognize the risk element. However, this kind of exposure netting (i.e. balancing of foreign currency positions on the basis of their co-movement) can be fully exploited by applying portfolio theory to the management of foreign exchange. The portfolio approach will allow the international company to determine correctly its overall exchange exposure, to build a well-diversified (minimum variance) currency portfolio, and to assess the impact of major investments on its company-wide foreign exchange exposure.

OBJECTIVE OF FOREIGN EXCHANGE MANAGEMENT

After an international company has consolidated the foreign exchange exposure of all subsidiaries and netted out assets and liabilities per currency, its net exposure will consist of long/short positions in various currencies. This set of foreign currency positions constitutes the

currency portfolio of the company. The objective of the international company's foreign exchange management is then to protect the company and its subsidiaries against exchange losses, but not to seek windfall gains through currency speculation or similar hedging activities unrelated to the nature of its business. The company's profitability should be based on the return on the goods and services it produces and sells, not on the return of its currency portfolio. Protection against currency risk is the objective, not speculative profits. The multinational company can reduce its foreign exchange risk by engaging in hedging transactions.

Hedging reduces the exposure and consequently the variance of the foreign currency portfolio, but also reduces its expected end-of-period value by the hedging costs incurred. In other words, the company should minimize the variance of its foreign currency portfolio given a maximum amount of hedging costs it is willing to incur, and not maximize the expected end-of-period return on the foreign exchange portfolio. Under efficient market conditions, only the contribution of the currency to the total risk of a portfolio should influence the value of the currency. All other risks (unsystematic risk) can be eliminated by holding a diversified portfolio of currencies. Unless the price movements of two currencies are perfectly positively correlated, any variation in the exchange rate of one that is not accompanied by an equivalent and simultaneous change in the rate of the other will lead to a reduced combined variation between both currencies. In other words, the combined variation of the two currencies will be less than the sum of the fluctuations in both exchange rates individually. The total risk of the currency portfolio is thus a function of the level of correlation among all currencies, the volatility of all currencies individually, and the proportion of each currency to the base currency equivalent sum of all currencies in the portfolio (i.e. the relative importance of each currency in the portfolio). In this portfolio-context, the objective of foreign exchange management is defined as the minimization of exchange risk to the company, i.e. to minimize the variance of the company's foreign currency portfolio subject to the costs incurred by hedging.

Hedging reduces variance caused by exchange rate changes but entails costs to the company. Therefore, one should concentrate on the relationship between the expected value and the variance of the company's foreign exchange portfolio at the end of the planning horizon. A portfolio model will now be developed to trace out the expected value-variance frontier for the company's foreign currency portfolio.

STATEMENT OF THE PROBLEM

The hedging problem can be formulated and solved as a decision problem. The data inputs required for the solution of the problem are: (a) a set of known variables (spot and forward exchange rates, domestic and foreign interest rates) and a set of variables assumed to be known (the foreign exchange exposure of the company): and (b) a set of unknown random variables (the future spot exchange rates for every currency in the portfolio at the end of the planning period). The decision variables are the amounts to be hedged in every currency in order to reduce the portfolio's variance to a selected level. These three sets of variables produce an uncertain value, V, i.e. the value of the company's foreign exchange portfolio at the end of the planning horizon, with expected value $\bar{V}$ and variance W. The hedging problem is then set up as the minimization of the variance of the portfolio subject to a set of operational constraints reflecting a specified maximum level of hedging costs and bounds on the amounts of hedging transactions.

Since the risk factor is expressed in quadratic terms, the solution technique has a quadratic objective function and hence requires quadratic programming. The model then chooses the amounts to be hedged in every currency to produce a desirable combination of expected value and variance. The flowchart in Fig. 1 illustrates the inputs and outputs of the model.

MODEL FORMULATION

1. *General notation*

The following notation will be used in the formulation of the model:

V = value of the firm at the end of the period.

Vol. 7
1979

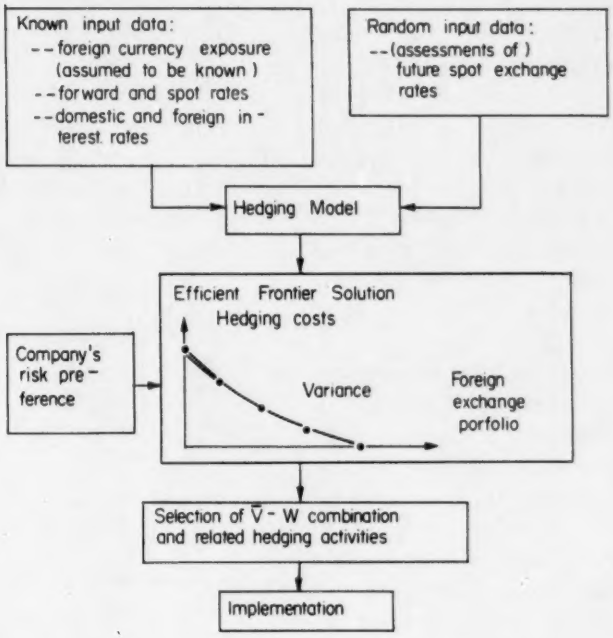


FIG. 1. *Inputs and outputs of the portfolio model.*

X_i = projected local currency denominated exposure in currency i at the end of the period.

i = subscript denoting the currencies in which the firm is conducting its business; i.e. $i = 1, 2, \ldots, N$; with $i = 1$ denoting the base currency of the company which is assumed to be the US dollar.

In order to be able to make a distinction between long and short positions in different currency exposures, we add the following conventional notation:

$i = 1$ for the base currency
$= 2, 3 \ldots, n$ for currencies in which the company has a *long* exposure
$= n + 1, n + 2, \ldots, N$ for currencies in which the company has a *short* exposure

$s_{0,i}$ = spot rate for currency i; $i = 2, \ldots, N$.

$s_{1,i}$ = future spot rate for currency i; $i = 2, \ldots, N$.

${}_0s_{1,i}$ = forward exchange rate for currency i; $i = 2, \ldots, N$.

$h_{1,i}$ = amount of exposure in currency i hedged in *the forward market*; $i = 2, \ldots, N$.

$h_{2,i}$ = amount of exposure in currency i hedged in *the Euro-currency market*; $i = 2, \ldots, N$.

$h_{3,i}$ = amount of exposure in currency i hedged in *the local money market*; $i = 2, \ldots, N$.

$[l_{1,i}; u_{1,i}]$ = lower and upper bounds of hedging transactions in the forward market for currency i.

$[l_{2,i}; u_{2,i}]$ = lower and upper bounds of hedging transactions in the Euro-currency market for currency i.

$[0; u_{3,i}]$ = lower and upper bounds of hedging transactions in the local money market for currency i.

TC = transaction costs per unit amount of hedging (i.e. brokerage fees, information costs and administrative costs).

TC_1 = unit TC in the forward market, expressed as a percentage of the hedging volume.

TC_2 = unit TC in the Euro-currency market, expressed as a percentage of the hedging volume.

TC_3 = unit TC in the local money market, expressed as a percentage of the hedging volume.

Note: We assume that all hedging transactions are taken at the beginning of the period. It is also assumed that there are no odd-day forward contracts and that all contracts are held to maturity.

r_i = Euro-currency interest rate for currency i; $i = 1, 2, \ldots, N$ with r_1 = Euro-dollar interest rate.

R_i = local money market interest rate for currency i; $i = 1, 2, \ldots, N$; with R_1 = US interest rate.

2. *The one-currency case*

To illustrate the derivation of the model, we consider the three hedging alternatives (forward market, Euro-currency market, local money market) consecutively and this for only *one* foreign currency (e.g. DM). Hence, X denotes the exposure in DM, assuming a long position ($X > 0$).

The end-of-period value of the firm (ignoring the fraction expressed in domestic currency, i.e. US \$) is then: $V = Xs_1$

Hedging in the forward market. The value of the foreign currency (DM) exposure at the end of the planning period as a function of the amount hedged can be represented as:

$$V = Xs_1 + (h_{1\,0}s_1 - h_1 s_1 - h_1 TC_1 s_0)$$

or,

$$V = Xs_1 + h_1({}_0s_1 - s_1 - TC_1 s_0)$$

Hedging in the Euro-currency market: Hedging a *long position* in a foreign currency consists of borrowing the foreign currency, converting these funds into the base currency (\$) and investing them in the money market at the going rate (i.e. *T*-bill rate).

$$V = Xs_1 + [h_2 s_0(1 + r_\$) - h_2(1 + r_{DM})s_1 - h_2 TC_2 s_0]$$
$$= Xs_1 + h_2[s_0(1 + r_\$) - (1 + r_{DM})s_1 - TC_2 s_0]$$

Similarly, hedging a *short position* in a foreign currency consists of borrowing the base currency (\$), converting them into local currency and investing the proceeds in the money market. Hence,

$$V = Xs_1 + [h_2 s_0(1 + r_{DM}) - h_2(1 + r_\$)s_1 - h_2 TC_2 s_0]$$
$$= Xs_1 + h_2[s_0(1 + r_{DM}) - s_1(1 + r_\$) - TC_2 s_0]$$

Assuming the long-DM exposure X can be covered by both hedging alternatives simultaneously, then:

$$V = [X - h_1 - (1 + r_{DM})h_2]s_1 + h_{1\,0}s_1 + h_2(1 + r_\$) - h_1 TC_1 s_0 - h_2 TC_2 s_0$$

Hedging in the local money market. The procedure is exactly similar to hedging in the Euro-currency market, only the notation changes, i.e. h_3, $R_\$$, R_{DM}, TC_3 instead of h_2, $r_\$$, r_{DM} and TC_2.

The V function is:

$$V = Xs_1 + [h_3 s_0(1 + R_{DM}) - h_3(1 + R_\$)s_1 - h_3 TC_3 s_0]$$
$$= Xs_1 + h_3[s_0(1 + R_{DM}) - s_1(1 + R_\$) - TC_3 s_0]$$

Considering the three hedging alternatives simultaneously, it can be shown that

$$V = Xs_1 - h_1[s_1 - {}_0s_1 + TC_1 s_0] - h_2[(1 + r_{DM})s_1 - (1 + r_\$)s_0 + TC_2 s_0] - h_3[(1 + R_{DM})s_1 - (1 + R_\$)s_0 + TC_3 s_0]$$

The expected value of V is the same, except that $\bar{s}$ needs to be substituted for s and the variance is

$$W = [X - h_1 - (1 + r_{DM})h_2 - (1 + R_{DM})h_3]^2 s_1(W)$$

Under the assumption of efficient markets and no transaction costs, however,

(a) $TC_1 = TC_2 = TC_3 = 0$

(b) $\bar{s}_1 = {}_0s_1$ (the forward rate is the unbiased forecast of the future spot rate)

(c) Interest Parity holds: $\bar{s}_1 = s_0\left(\frac{1 + r_\$}{1 + r_{DM}}\right)$

$\rightarrow \bar{s}_1(1 + r_{DM}) = s_0(1 + r_\$)$

and similarly

$\bar{s}_1(1 + R_{DM}) = s_0(1 + R_\$)$

Vol. 7
1979

The 'HEDGE' strategy leads to:

$$V = Xs_1$$

$$\bar{V} = X\bar{s}_1 = X_0 s_1$$

$$W = [X - h_1 - (1 + r_{DM})h_2 - (1 + R_{DM})h_3]^2 s_1(W)$$

The 'NO HEDGE' strategy implies:

$$V = Xs_1$$

$$\bar{V} = X_0 s_1$$

$$W = X^2 s_1(W)$$

Comparing the 'HEDGE' with the 'NO HEDGE' policy under the assumption of efficient markets and no transaction costs; it can be deduced that the right policy is to hedge all exposure all the time. The expected value of the firm is the same, but the variance of the firm's end-of-period value is reduced under the 'HEDGE' strategy. *The N-currency case* can be similarly developed and, again, under efficient market conditions and zero transaction costs,

(a) $TC_1 = TC_2 = TC_3 = 0$

(b) $\bar{s}_{1,i} = {}_0 s_{1,i}$

(c) $\begin{cases} s_{0,i}(1 + r_1) = \bar{s}_{1,i}(1 + r_i) \\ s_{0,i}(1 + R_1) = \bar{s}_{1,i}(1 + R_i) \end{cases}$

(d) $\begin{cases} s_{0,i}(1 + r_i) = \bar{s}_{1,i}(1 + r_1) \\ s_{0,i}(1 + R_i) = \bar{s}_{1,i}(1 + R_1). \end{cases}$

The 'HEDGE' strategy implies:

$$V = X_1 + \sum_{i=2}^{N} X_i s_{1,i}$$

$$\bar{V} = X_1 + \sum_{i=2}^{N} X_{i0} s_{1,i}$$

W = same expression as before

Under the assumption of efficient markets and no transaction costs, ex-ante costs of hedging are zero, i.e. one can reduce the variance W to zero at no costs. However, ex-post hedging costs might differ from zero, but will still be relatively small since deviations occur on both sides of the Interest Parity line. In practice, transaction costs are also non-zero. Figure 2 represents the relation between the expected value of the firm and the variance of this value for both ex-ante and ex-post hedging costs under the efficient market hypothesis. The 'NO HEDGE' strategy implies:

$$V = X_1 + \sum_{i=2}^{N} X_i s_{1,i}$$

$$\bar{V} = X_1 + \sum_{i=2}^{N} X_{i0} s_{1,i}$$

$$W = \sum_{i=2}^{N} \sum_{j=2}^{N} X_{i0} s_{1,i} X_{j0} s_{1,j} \mathrm{Cov}(s_{1,i}; s_{1,j}).$$

Hence, under the efficient market hypothesis and no transaction costs, the recommended strategy is to hedge all exposed currency positions, because compared with the 'no hedge' strategy, the variance of the value of the firm at the end of the period is lower while the expected value remains unchanged.

CONCLUSIONS

The portfolio approach to foreign exchange exposure management makes an explicit consideration of the inherent relationships among the currencies in the company's foreign exchange portfolio. It also considers the costs of hedging per individual currency and this for three alternative methods of hedging (i.e.

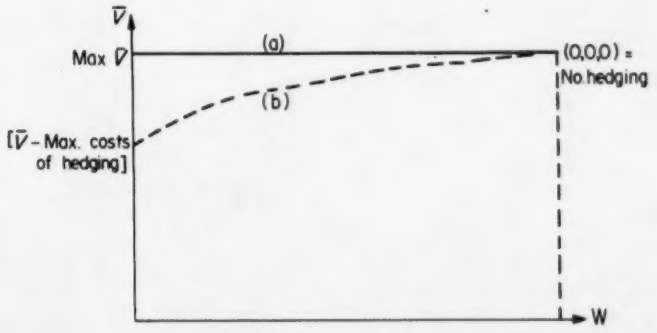


FIG. 2. *(a) Ex-ante costs of hedging and zero transaction costs. (b) Ex-post costs of hedging and positive transaction costs.*

forward market, Euro-currency market, local money market).

The quadratic programming model traces out an 'efficient frontier' between the variance of the company's currency portfolio and the expected value of this portfolio at the end of the planning period. For every point on the frontier the program selects the optimal mix of hedging activities, i.e. the amount of exposure to be hedged and the method of hedging to be used. In doing so, the model chooses the optimal amount and method of hedging in each currency, minimizing the variance of the company's foreign exchange portfolio, corresponding to a selected level of total hedging costs (end-of-period expected value of the company's foreign currency portfolio) to be incurred.

ACKNOWLEDGEMENTS

The author wishes to thank Professor W White, R Glauber, and P Jones of the Harvard University Graduate School of Business Administration for their advice and assistance in the course of this study. Of course, I alone bear full responsibility for its content and shortcomings.

ADDRESS FOR CORRESPONDENCE: *LA Soenen, Assistant Professor, Department of Industrial Engineering and Management Science, Eindhoven University of Technology, 5600 MB, Eindhoven, The Netherlands.*

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci. Vol. 7, No. 4, pp. 345 to 351
© Pergamon Press Ltd 1979. Printed in Great Britain

0306-0483/79/0801-0345$02.00/0

Allocation of Research and Development Funds: A Zero-One Goal Programming Approach

ARTHUR J KEOWN

BERNARD W TAYLOR III

Virginia Polytechnic Institute and State University

CALVIN P DUNCAN

Washington State University

(*Received November 1977; in revised form January 1979*)

The research and development project selection process is one of the most difficult and important problems faced by management. It is typically complicated by indivisibility of projects and multiple and conflicting objectives, in addition to limitations on funding, facilities, and qualified researchers. In this paper a case example involving a high technology electrical equipment manufacturer is developed to illustrate this problem using zero–one goal programming to accommodate indivisibility of projects in addition to multiple and conflicting goals. The model presented is an attempt to provide managers with a robust tool for allocating scarce resources among research and development projects.

INTRODUCTION

RESEARCH and development operations are vital to the growth of organizations in competitive markets. Decisions concerning these projects at their embryo stage can dictate the eventual success or failure of a firm years later. While the selection of research and development projects is one of the most important decisions facing the firm, it is also one of the most difficult. In making this decision, managers are faced with the problem of uncertain outcomes compounded by indivisibility of projects and multiple and conflicting objectives, in addition to limitations on funding, facilities, and qualified researchers.

The existence of multiple and conflicting objectives has received wide support in the literature, particularly in the R&D project selection process. In reviewing the literature, Baker [2, p. 166] concludes that "it is clear that the R&D project selection decision can be characterized as having multiple criteria with no common, underlying measure." Thus, the major stumbling block in developing an R&D project selection model is the need for the model to reflect these multiple goals. This paper presents a description and application of a zero–one goal programming model to the R&D project selection problem, thus allowing for goal satisfaction in a sequential and/or simultaneous manner.

While the R&D selection problem has produced numerous articles advocating various methodologies, it has yet to be adequately solved from the point of view of the R&D manager. This is reflected by the few quantitative R&D project selection models that have been actually implemented and used [2]. Scoring models, see [4], have received considerable attention in the literature. These models determine an overall benefit measure by combining various project scores from numerous decision criteria, either through addition or multiplication. However, the conversion of a multidimensional vector of goals into a unidimensional goal is by its very nature difficult when

these goals are measured in incommensurable units. For this reason this approach is not a meaningful alternative.

An alternative approach to the problem of multiple conflicting goals where these goals are measured in incommensurable units is goal programming, see for example [7]. Instead of forcing the objective function to take on a unidimensional character, goal programming through a hierarchical optimization procedure allows the various goals to be scalar weighted and/or ordinally ranked. Flexibility is thereby provided in the determination of an optimal solution for management to reflect its individual preferences and policies in a lexicographic manner (i.e. the establishment of a system of priorities).

While goal programming provides a logical approach to the problem of multiple and conflicting objectives the R&D project selection problem is further complicated by the indivisibility of projects. This leads naturally to a combination of zero–one integer programming and goal programming. Moreover, the use of zero–one goal programming enables the R&D manager to deal with the complexities of reality. The introduction of integer programming to the model formulation allows for the inclusion of mutually exclusive projects, and contingent projects, in addition to requiring integer as opposed to fractional solutions. As each of these situations is commonplace in capital budgeting problems, their exclusion from the problem could lead to suboptimization.

ZERO–ONE GOAL PROGRAMMING APPLICATION

A case study, based upon a high technology electrical equipment manufacturer will be presented to illustrate the use of zero–one goal programming. Because the objectives and goals of any research and development operation must reflect the specific goals of the parent organization, no one goal formulation should be expected to fit every situation. Thus, the model formulation presented herein is just one of the many alternative goal combinations possible.

The problem for the company is posed by the availability of sixteen possible R&D projects. The relevant information concerning these projects is presented in Table 1. This firm is composed of two major divisions—the Atomic, Defense, and Space Division and the Electric Utility Division—and because of its size is faced with limited R&D funds in addition to limited resources with respect to both laboratories and qualified researchers in both divisions. The firm has two major research centers, one in the Mid Atlantic Region specializing in electrical utility research and the other in the Southeast, specializing in atomic, defense, and space research. While some research can be accommodated at either location, most utility research is limited to the Mid Atlantic center, and most atomic, defense, and space research is limited to the Southeastern center.

VARIABLES, CONSTRAINTS AND GOALS

Decision variables

The decision variables in the model include R&D projects being considered for possible adoption. Since fractional acceptance of these projects is unacceptable they are expressed as zero–one values. In addition, the objective function will involve the minimization of deviational variables from the various goals where these deviational variables are scalar weighted and/or ordinally ranked. Thus the relevant decision variables are:

$X_i = \begin{cases} 0 \text{ if project } i \text{ is not accepted } (i = 1, \ldots, m) \\ 1 \text{ if project } i \text{ is accepted} \end{cases}$

d^- = The negative slack deviation variable

d^+ = The positive slack deviation variable

Strict constraints

The use of the zero–one goal programming approach allows for the inclusion of mutually exclusive projects, in addition to requiring integer as opposed to fractional solutions. Projects X_9 and X_{10} are mutually exclusive because they both require the use of the same specialized laboratory equipment. In addition, the research on three projects can be performed at either of the two research facilities. Thus, because of the limited research facilities, these projects are included twice in the model, with their research being performed at each facility. They are X_1 and X_2, X_7 and X_8, and X_{19} and X_{20}. These mutually exclusive projects are represented by the following equations:

$$X_1 + X_2 \leqslant 1.0$$
$$X_7 + X_8 \leqslant 1.0$$
$$X_9 + X_{10} \leqslant 1.0$$
$$X_{19} + X_{20} \leqslant 1.0$$

Vol. 7
1979

TABLE I.

	Project description	Cost ($)	Physical facilities research center I (%)	Physical facilities research center (II) (%)	Manpower usage research center I (men)	Manpower usage research center II (men)	Offensive (O) Defensive (D) nature of project	Sucess/ proba-bility (%)	Sales–atomic, defense, and space division ($)	Sales–electric utility division ($)	Market share atomic equip. market (%)	Steam-powered electric utility market share (%)	Net present value ($)
X_1	Improvement on main coolant pumps—atomic energy plants	$220,000	18%		5		D	30%	$420,000		2.67		$180,000
X_2	Improvement on main coolant pumps—atomic energy plants	220,000		15%		5	D	30%	420,000		2.67		180,000
X_3	New control drive mechanism—atomic energy operations	90,000	14%		7		O	18%	250,000		0.50		40,000
X_4	Aerospace power conditioning & conversion equipment	300,000	24%		9		O	10%	1,500,000				70,000
X_5	Thermoelectric cooling device for atomic military applications	390,000	28%		9		O	15%	1,233,000		8.00		100,000
X_6	Nuclear reactor core components for naval applications	310,000	35%		5		O	23%	674,000		3.47		130,000
X_7	Improved pump for controlled circulation boilers on atomic energy plts.	280,000	23%		6		D	35%	428,000		2.857		80,000
X_8	Improved pump for controlled circulation boiler on atomic energy plants	280,000		20%		6	D	35%	428,000		2.857		80,000
X_9	Improved tracking & discrimination system	140,000	20%		2		D	30%	223,000				110,000
X_{10}	Guidance & control system for space veh.	340,000	23%		4		D	25%	660,000				130,000
X_{11}	Anti-missile protection system	320,000	21%		5		O	20%	750,000				60,000
X_{12}	Improved surface condenser-steam division	170,000		25%		3	D	30%		$300,000		2.00	40,000
X_{13}	Improved priming ejector	220,000		20%		8	D	33%		303,000		2.27	110,000
X_{14}	Improved water conversion plant—for installation on land	200,000		18%		4	D	30%		333,000		2.33	70,000
X_{15}	Improved flash evaporator	100,000		14%		4	D	33%		136,000		1.12	30,000
X_{16}	New high voltage circuit breaker	300,000		19%		2	O	18%		833,000			80,00
X_{17}	New dry & liquid immersed distribution transformer	330,000		28%		7	O	10%		1,550,000			70,000
X_{18}	Printed circuit modules	250,000		23%		6	O	18%		666,000			110,000
X_{19}	Improved nuclear steam generator	420,000	31%		8		D	40%		550,000		3.50	140,000
X_{20}	Improved nuclear steam generator	420,000		31%		8	D	40%		550,000		3.50	140,000

(In a similar manner, if the acceptance of project X_k is dependent upon the prior acceptance of project X_j perhaps because project X_k is an extension of project X_j, $X_k - X_j \leqslant 0.0$.)

Goals

R&D budget goal. The level of resources available for R&D projects is generally limited by a budget constraint; in this study, the R&D budget goal provides for an expenditure ceiling of $2,100,000 on new R&D projects, i.e.

$$\begin{aligned}&\$220{,}000X_1 + \$220{,}000X_2 + \$90{,}000X_3 + \$300{,}000X_4\\&+ \$390{,}000X_5 + \$310{,}000X_6 + \$280{,}000X_7\\&+ \$280{,}000X_8 + \$140{,}000X_9 + \$340{,}000X_{10}\\&+ \$320{,}000X_{11} + \$170{,}000X_{12} + \$220{,}000X_{13}\\&+ \$200{,}000X_{14} + \$100{,}000X_{15} + \$300{,}000X_{16}\\&+ \$330{,}000X_{17} + \$250{,}000X_{18} + \$420{,}000X_{19}\\&+ \$420{,}000X_{20} - d_1^+ \leqslant \$2{,}100{,}000.\end{aligned}$$

This goal is achieved through the minimization of d_1^+.

Physical facilities goal. Not only are R&D activities limited by constraints on funding, but also by limited research facilities. If, for example, the research activities associated with specific subsets of the projects being considered must be performed at specific research laboratories or involve the use of specific research machinery for certain lengths of time, then the project acceptance process will be constrained by the total capacity of the research laboratories or machinery. The number of physical facilities constraints that exist will reflect the number of limited facilities for which the R&D projects are vying.

In the case study, it was desired that the new projects being selected do not require more in the way of physical research facilities than are available at the firm's two research centers.

$$\begin{aligned}&0.18X_1 + 0.14X_3 + 0.24X_4 + 0.28X_5 + 0.35X_6\\&+ 0.23X_7 + 0.20X_9 + 0.23X_{10} + 0.21X_{11}\\&+ 0.31X_{19} - d_2^+ \leqslant 1.00\\&0.15X_2 + 0.20X_8 + 0.25X_{12} + 0.20X_{13} + 0.18X_{14}\\&+ 0.14X_{15} + 0.19X_{16} + 0.28X_{17} + 0.23X_{18}\\&+ 0.31X_{20} - d_3^+ \leqslant 1.00.\end{aligned}$$

This goal is achieved through the minimization of d_2^+ and d_3^+.

Maximum manpower goals. The unavailability of qualified researchers may also force the firm to further limit the acceptance of R&D projects. Due to the shortage of talented researchers, the firm may wish to establish an upper limit on the number of research personnel needed to manage new R&D projects. Alternatively the management might specify some absolute limit on R&D project expenditures which can be adequately managed with the available research manpower.

For a particular firm, if the research personnel are not of a homogeneous nature, that is, different projects require different types or categories of researchers, then there would be as many of these goals as there are different categories of researchers. Each goal would limit the manpower needed in a particular research category created by the acceptance of new projects to the maximum available research personnel for the category.

In the study, the management wishes to ensure that the new projects being selected do not require more in the way of qualified researchers than are available at the two research centers. There were 22 researchers available at research center I and 25 researchers available at research center II.

$$\begin{aligned}&5X_1 + 7X_3 + 9X_4 + 9X_5 + 5X_6 + 6X_7\\&\qquad + 2X_9 + 4X_{10} + 5X_{11} + 8X_{19} - d_4^+ \leqslant 22\\&5X_2 + 6X_8 + 3X_{12} + 8X_{13} + 4X_{14} + 4X_{15}\\&\qquad + 2X_{16} + 7X_{17} + 6X_{18} + 8X_{20} - d_5^+ \leqslant 25.\end{aligned}$$

Achievement of this goal is obtained through the minimization of the positive deviational variables.

Priority projects goal. The management of the firm may have certain projects which it feels are of utmost importance. This being the case, it may feel it desirable to accept at least one or more projects from some subset of projects.

In this example, management is quite concerned with improving its position in the space and defense areas and thus wants at least two projects selected from the set of projects X_5, X_6, X_9, X_{10}, and X_{11}.

$$X_5 + X_6 + X_9 + X_{10} + X_{11} + d_6^- \geqslant 2$$

The minimization of d_6^- reflects this goal.

Offensive–defensive project balance goal. R&D projects are often categorized as being either offensive, concerned with new project development, or defensive, concerned with im-

Vol. 7
1979

provements in existing projects. In general, firms attempt to achieve a reasonable balance between these categories of projects in order to ensure that current products are periodically upgraded and to ensure the development of new products. This goal can be represented by forcing a minimum dollar amount to be spent in each category or providing for the acceptance of a minimum number of projects from each category.

In this situation, the firm wishes projects to be spread between defensive and offensive projects. Specifically, they wish at least five defensive and three offensive projects to be adopted.

$$X_3 + X_4 + X_5 + X_6 + X_{11} + X_{16} + X_{17} + X_{18} + d_7^- \geq 3.0$$

$$X_1 + X_2 + X_7 + X_8 + X_9 + X_{10} + X_{12} + X_{13} + X_{14} + X_{15} + X_{19} + X_{20} + d_8^- \geq 5.0.$$

Negative deviational variables, d_7^- and d_8^-, are minimized in this goal.

Risk spreading goal. Because the probability of success on R&D projects varies greatly, it may be desirable to ensure that all projects selected are not of the highest risk level, or alternatively, to provide for some minimum level of expected success in the set of projects selected. This could be implemented either by categorizing projects by risk class and simply limiting the number of projects selected in the higher risk level or setting a minimum on the number of projects being selected in the lower categories. Alternatively, the firm could set a minimum for the sum of probabilities of success of projects being selected. In either approach the firm would force the acceptance of some of the less risky projects.

The policy adopted by management in this study requires the sum of the probabilities of success of new projects being selected to be at least 2.00 or more.

$$0.30X_1 + 0.30X_2 + 0.18X_3 + 0.10X_4 + 0.15X_5 + 0.23X_6 + 0.35X_7 + 0.35X_8 + 0.30X_9 + 0.25X_{10} + 0.20X_{11} + 0.30X_{12} + 0.33X_{13} + 0.30X_{14} + 0.33X_{15} + 0.18X_{16} + 0.10X_{17} + 0.18X_{18} + 0.40X_{19} + 0.40X_{20} + d_9^- \geq 2.00.$$

Goal achievement is reflected by the minimization of d_9^-.

Sales goal. The firm may desire that the increase in expected average annual sales over the 10-year period following the completion of the selected projects be greater than or equal to some level. This desired sales increase may be for the firm as a whole, for each division, or for particular product markets.

In the illustrative case the firm wishes the increase in expected average annual sales from the acceptance of new projects in the atomic, defense, and space division and electric utility division for the 10-year period after the completion of the projects to be \$750,000 and \$500,000 respectively.

$$0.3(\$420{,}000)X_1 + 0.3(\$420{,}000)X_2 + 0.18(\$250{,}000)X_3 + 0.10(\$1{,}500{,}000)X_4 + 0.15(\$1{,}233{,}000)X_5 + 0.23(\$674{,}000)X_6 + 0.35(\$428{,}000)X_7 + 0.35(\$428{,}000)X_8 + 0.30(\$223{,}000)X_9 + 0.24(\$660{,}000)X_{10} + 0.20(\$750{,}000)X_{11} + d_{10}^- \geq \$750{,}000$$

$$0.30(\$300{,}000)X_{12} + 0.33(\$303{,}000)X_{13} + 0.30(\$333{,}000)X_{14} + 0.33(\$136{,}000)X_{15} + 0.18(\$833{,}000)X_{16} + 0.10(\$1{,}550{,}000)X_{17} + 0.18(\$666{,}000)X_{18} + 0.40(\$550{,}000)X_{19} + 0.40(\$550{,}000)X_{20} + d_{11}^- \geq \$500{,}000.$$

The achievement of these goals is accomplished by minimizing d_{10}^- and d_{11}^-.

Market share growth goals. The firm may also wish to secure an expected average annual increase in their share of one or more particular markets. The existence of a market share growth goal draws support from the Kaplan, Dirlan and Lanzillotti [6] study, in particular the Pacific Tea Company, Swift and Company, and Standard Oil Company cases. There will be as many of these goals as there are markets in which the firm desires to increase its share.

The firm wishes to secure an expected average annual increase in market share over the 10-year period following the completion of the projects in the atomic equipment and steam-powered electric utility markets of 2.8 and 2.2% respectively.

$$0.30(2.67)X_1 + 0.30(2.67)X_2 + 0.18(0.50)X_3 + 0.15(8.0)X_5 + 0.23(3.47)X_6 + 0.35(2.857)X_7 + 0.35(2.857)X_8 + d_{12}^- \geq 2.80$$

$$0.30(2.00)X_{12} + 0.33(2.27)X_{13} + 0.30(2.33)X_{14} + 0.33(1.12)X_{15} + 0.40(3.50)X_{19} + 0.40(3.50)X_{20} + d_{13}^- \geq 2.20.$$

Goal achievement is represented by the minimization of d_{12}^- and d_{13}^-.

Maximization of net present values. The management also desires to maximize the net present value of the set of R&D projects it selects. This is done through the minimization of d_{14}^-.

$$\$180{,}000X_1 + \$180{,}000X_2 + \$40{,}000X_3 + \$70{,}000X_4 + \$100{,}000X_5 + \$130{,}000X_6 + \$80{,}000X_7 + \$80{,}000X_8 + \$110{,}000X_9 + \$130{,}000X_{10} + \$60{,}000X_{11} + \$40{,}000X_{12} + \$110{,}000X_{13} + \$70{,}000X_{14} + \$30{,}000X_{15} + \$80{,}000X_{16} + \$70{,}000X_{17} + \$110{,}000X_{18} + \$140{,}000X_{19} + \$140{,}000X_{20} + d_{14}^- \geq \$10{,}000{,}000.$$

OBJECTIVE FUNCTION

The objective function for the R&D project selection problem involves the minimization of the appropriate deviations which are weighted and/or ranked in a manner dictated by the firm's management. A company, depending upon its particular environment and goals, may desire a unique model formulation and subsequent objective function. Fortunately, the zero–one goal programming technique is robust enough to accommodate unlimited variations.

In this study the firm's management has developed the following priority structure for the R&D project selection model:

Priority 1: R&D Budget Goal
Priority 2: Physical Facilities Goals
Priority 3: Maximum Manpower Goals
Priority 4: Priority Projects Goal
Priority 5: Offensive–Defensive Project Balance Goal
Priority 6: Risk Spreading Goal
Priority 7: Sales Goal
Priority 8: Market Share Growth Goals
Priority 9: Maximization of Net Present Value.

This priority structure results in the following objective function:

$$\text{Minimize } Z = P_1 d_1^+ + P_2[d_2^+ + d_3^+] + P_3[d_4^+ + d_5^+] + P_4 d_6^- + P_5[d_7^- + d_8^-] + P_6 d_9^- + P_7[d_{10}^- + d_{11}^-] + P_8[d_{12}^- + d_{13}^-] + P_9 d_{14}^-.$$

The priority structure and subsequent objective function formulation evolves from the firm's preferences and desires with respect to the selection of new R&D projects. In addition, the use of preemptive priorities in the objective function reflects the firm's desire to have the various goals treated in a lexicographic manner. As this model formulation reflects the individual preferences and desires of one particular firm, it should not be construed as the only possible valid model formulation. Each firm, because of its particular decision environment, will have its own individual model formulation and goal priority structure. Thus, the objective function must be viewed as being relevant only to the firm for which it applies.

MODEL RESULTS

In total, the zero–one goal programming model involved twenty variables, five system constraints, fourteen goal equations, and an objective function. The model was solved by the Lee and Morris algorithm, based upon the implicit enumeration method of Balas [3] and the backtracking technique of Glover [5]. The solution to this problem involved the acceptance of eight projects:

X_1 = Improvement on main coolant pumps—atomic energy plants
X_4 = Aerospace power conditioning and conversion equipment
X_6 = Nuclear reactor core components for naval applications
X_8 = Inproved pump for controlled circulation boiler on atomic energy plants
X_9 = Improved tracking and discrimination system
X_{12} = Improved surface condenser–steam division
X_{18} = Printed circuit modules
X_{20} = Improved nuclear steam generator

The model completely satisfied the first six goals (P_1 through P_6); however, it failed to satisfy goals P_7 through P_9. The degree of underachievement experienced was as follows:

Priority 7: Underachievement of \$172,400. The desired increase in expected average annual sales from the acceptance of new projects in the atomic, defense, and space division of \$75,000 fell short by \$102,280, and the desired increase in expected average sales from the acceptance of new projects in the electric utility division of \$500,000 fell short by \$70,120.

Priority 8: Underachievement of 0.401%. The desired expected annual increase over the ten-year period following the completion of the projects in the atomic equipment and steam

Vol. 7
1979

powered electric utility market of 2.8 and 3.0% respectively from the acceptance of new projects fell short by 0.20095% in the atomic equipment market and 0.2% in the steam powered electric utility market.

Priority 9: The net present value of the projects selected was $860,000.

IMPLEMENTATION

The output of the zero–one goal programming model was used by the management as an additional decision making tool rather than a strictly binding decision. As a result the ultimate selection of projects by management produced a set of projects that closely paralleled those identified by the zero–one goal programming model. The use of the model and its subsequent results were viewed favorably by management in that it forced them to define and establish goals, required a quantification of the resulting benefits, and provided a guideline in the selection process.

CONCLUSION

The research and development project selection process is one of the most difficult and important problems faced by management. It is typically characterized by indivisibility of projects, multiple and conflicting objectives, measured in incommensurable units, and resource limitations on finding research facilities and qualified researchers. Because of the importance of these constraints and complications, they must be included in the modeling process in order to provide meaningful results. The zero–one goal programming approach provided here allows for each complication. This model has been an attempt to provide managers with a robust tool for allocating scarce resources among research and development projects.

REFERENCES

1. Baker NR (1974) R&D project selection models: an assessment. *IEEE Trans. Engng Mgmt*, EM-21, 165–171.
2. Baker NR & Pound WH (1964) R&D project selection: where we stand. *IEEE Trans. Engng Mgmt*, EM-11, 124–134.
3. Balas E (1965) An additive algorithm for solving linear programs with zero-one variables. *Ops Res.* 13, 517–545.
4. Dean BV (1970) *Project evaluation: methods and procedures. Am. Mgmt Rep.* 39, New York.
5. Glover F (1965) A multiphase–dual algorithm for zero–one integer programming problem. *Ops Res.* 13, 879–919.
6. Kaplan ADH, Dirlan JB & Lanzillotti (1960) *Pricing in Big Business: a Case Approach.* Brookings Institute, Washington DC, USA.
7. Lee Sang M (1976) *Linear Optimization for Management.* Appendix III. Petrocelli/Charter, New York, USA.

ADDRESS FOR CORRESPONDENCE: *Arthur J Keown, Esq, Associate Professor of Finance, Virginia Polytechnic Institute and State University, Department of Business Administration, Blacksburg, Virginia* 24061, *USA*

BLANK
PAGE

Vol. 7
1979

MEMORANDA

ol. 7
979

BLANK
PAGE

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 4, pp. 355 to 374
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0801-0355$02.00/0

An Exercise In Probability Encoding

THE PURPOSE OF ENCODING

DECISION theory has been developed as a systematic methodology for decision-makers to apply in weighing alternative courses of action in the face of uncertainty [2]. Apart from mapping out a set of events branching from nodes in a decision tree, the decision-maker needs to be explicit about two issues: first, about his attitude to risk, and secondly about the probability that a given event or outcome in the tree will materialise. Surveys on the use of decision theory in practice suggest that one of the problem areas encountered by practitioners is the assessment of probabilities and probability distributions (see, for example, [1]). Proponents of decision theory recognise this difficulty, and typically echo the argument of Staël von Holstein [4]: "There is substantial evidence that people do not respond to uncertainty in perfect consistency with the axioms of probability. Sometimes it is also clear that the responses do not agree with the underlying judgement, even though they may be internally consistent. It is therefore important to develop efficient procedures for probability encoding, that is, the process of extracting and quantifying individual judgement about uncertain quantities."

Numerous psychological experiments to assess man's ability to act as an 'intuitive statistician' and to estimate probability values have been carried out, and some references are listed in [3]. Most of these experiments have been carried out in a laboratory environment and their results may therefore be regarded by some as being of very limited value. However, it is rather doubtful whether data collection can be properly carried out for decision situations in real life; not only would it be difficult to compare actual outcomes with prior probability estimates when circumstances following decisions change prior to the outcomes being ascertained, but the inevitable passage of time during which observations are recorded leads to changes in the decision-makers themselves, so that their responses to stimuli cannot then be aggregated to simple statements about their ability to estimate probabilities.

Furthermore, it is clear that when an analyst cross-examines decision-makers for the purpose of determining probability distributions, various motivational and cognitive biases play a part [3]. As stated earlier, encoding procedures are designed for the purpose of extracting information about an individual assessment of probabilities, values or both [3]:

"*P*-methods require the subject to respond by specifying points on the probability scale while the values remain fixed.

V-methods require the subject to respond by specifying points on the value scale while the probabilities remain fixed.

PV-methods ask questions that must be answered on both scales jointly; the subject essentially describes points on the cumulative distribution".

Of these, it is arguable that *P*-methods are less susceptible to motivational bias than methods that require assessments of values, and it therefore follows that laboratory-type experiments that concentrate on the assessment of probabilities are better suited for an examination of people's cognitive ability than observations of decision-makers in action.

THE PROBABILITY WHEEL

How are probability values determined? A popular encoding technique involves the use of a probability wheel, described by Spetzler and Staël von Holstein [3]:

> "The probability wheel is one of the most useful tools we have discovered for encoding indirect responses from subjects. The wheel is a disk with two adjustable sectors, one blue and the other orange, with a fixed pointer in the middle. ... A simple adjustment changes the relative size of the two sectors and thereby also the probabilities of the pointer indicating either sector when the disk stops spinning. The subject is asked which of the two events he considers more likely—the event relating to the uncertain quantity (for example, the event that new year's production will not exceed x units), or the event that the pointer ends up in the orange sector. The amount of orange in the wheel is then varied until the subject finds the two events equally likely. The relative amount of orange is then assigned as the probability of the event".

It is arguable whether such an encoding technique requires us to presuppose that an individual with a highly developed cognitive ability in comparing two events (the one relating to the matter under discussion and the other associated with the wheel in front of him) must also be capable of a quantitative assessment of the relative magnitude of the two sectors on the wheel. There are two reasons for this assumption: first, it would be difficult to explain how an individual who is poor in estimating the relative size of areas can be good at the much more complicated task of assessing probabilities of future events (and it would certainly be difficult to test such a contention); secondly, the ability to estimate relative areas can be objectively tested, whereas probabilities of future events must depend on subjective assessment by the individuals in question.

AN EXPERIMENT AND SOME RESULTS

Twenty slides were prepared, each showing an area divided into two sectors, one hatched and the other blank (see Fig. 1). The slides were shown, one at a time (and in the order presented in Fig. 1) to 62 individuals (all post-graduate students in management science, some with industrial experience), asking them to estimate the percentage of the hatched areas.

The results are summarised in the histograms in Fig. 2, where deviations from the actual values are shown, and in Table 1, giving the number of overestimators and underestimators as well as the average error for each group.

These results suggest a wide spectrum of variability, not only in the range of values covered by the histograms, but also in the preponderance of overestimators in some cases contrasted by the preponderance of

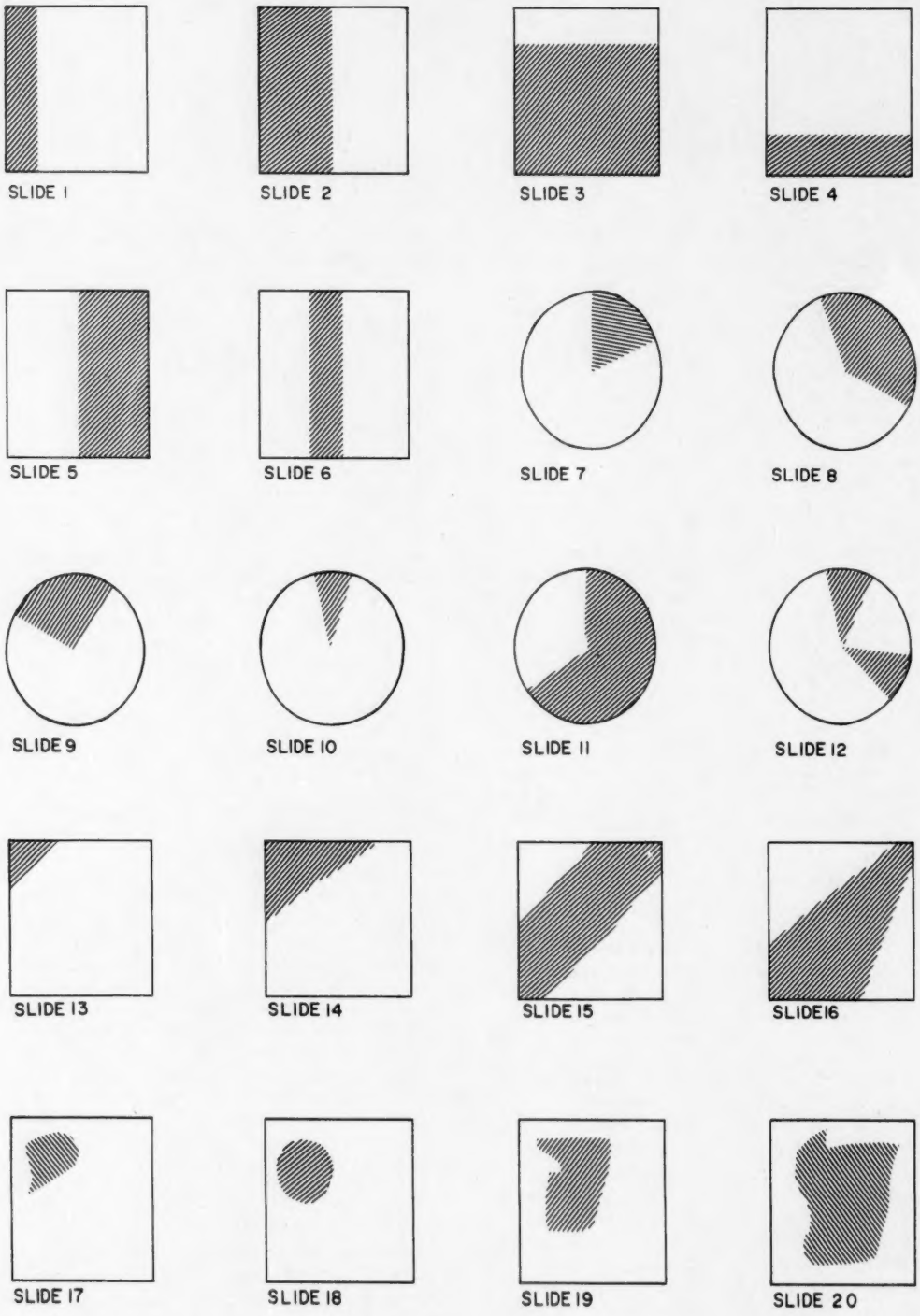


FIG. 1. 20 *slides for estimating the hatched area*

underestimators in others. The cognitive ability of an individual may well be affected by the shape of the area shown in the slide, but the actual value of the hatched area appears to be a dominant factor as well: most of the estimates are multiples of five and peaks in the histograms coincide with these points (they are highlighted by arrows shown on the histograms; for example, in slide no. 1 the actual value is 22%, so that the nearest multiples of five lie at 20% and 25%, involving estimating errors of -2 and $+3$ respectively, as shown on the histogram for slide 1). Hence slides 2 and 4, for which the actual value is 50%, generated relatively few errors, and similarly other slides with actual values which are multiples of five had a fairly high number of correct estimates, although in some cases their number was almost matched or even

Vol. 7
1979

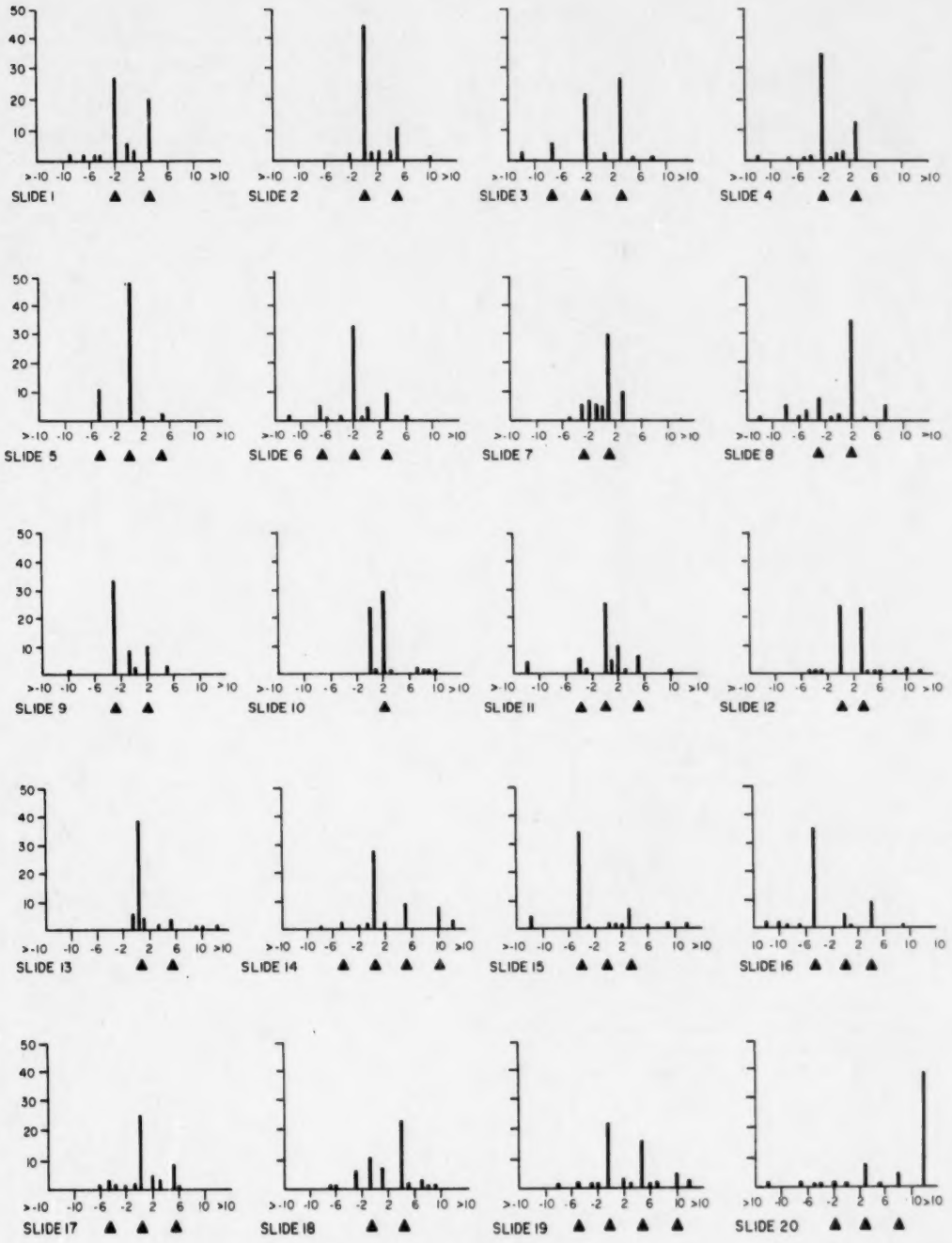


FIG. 2. *Histograms of results of 62 respondents*
(Notes: the abscissa shows the absolute error in estimating, the ordinate gives the frequencies, and the arrows indicate multiples of 5 in the estimates)

exceeded by the number of overestimators (as in slides 14, 17 and 20).

Only five slides (Nos. 7–11) resemble the probability wheel, while a sixth slide (No. 12) has two hatched segments (instead of one) in a disk. The results do not suggest that the cognitive ability of subjects faced with the disks whown in these slides was superior to that when faced with rectangular segments hatched in square areas (as in slides 1–6). The preponderance of those estimating in multiples of five for slides 7–12 is again evident in the histograms in Fig. 2.

Users of the probability wheel and similar devices in encoding procedures should therefore be aware not only of the difficulties that some decision-makers find in maintaining consistency in their estimates, but also of the bias which may creep in, caused by the device itself and/or by the numerical value of the probability being estimated.

TABLE 1. SUMMARY OF RESULTS

Slide No.	Shaded area (%)	Number of			Average number of	
		Under estimates	Correct estimates	Over estimates	Under estimates	Over estimates
1	22	32	5	25	2.5	2.7
2	50	1	44	17	1.0	4.5
3	77	31	0	31	8.5	3.1
4	22	42	3	17	2.9	2.5
5	50	11	48	3	5.0	4.0
6	22	45	5	12	3.3	3.5
7[1]	18.5[2]	20	5	36	2.2	1.7
8	38	19	2	41	5.4	2.9
9	28	46	3	15	2.8	2.3
10	7.6[2]	0	23	39	0	2.8
11	64.6[2]	13	25	24	9.5	3.0
12	21.6[2]	28	3	30	1.5	4.4
13	4.5[2]	8	39	15	1.1	9.9
14	20	7	28	27	3.9	9.9
15[1]	55.5[2]	42	2	17	6.3	5.6
16	55.5[2]	44	6	12	6.3	4.2
17	10	13	26	23	3.6	4.2
18	11	21	0	41	2.2	3.8
19	20	6	22	34	4.3	6.3
20	37	8	1	53	6.4	13.2

Notes:
1. In two cases only 61 participants responded.
2. Estimates to the nearest integer above or below were regarded as involving 0% error.

ACKNOWLEDGEMENT

The help of Miss J Everett in compiling the results is gratefully acknowledged.

REFERENCES

1. LONGBOTTOM D & WADE G (1973) An investigation into the application of decision analysis in United Kingdom companies. *Omega* **1**(2), 207–215.
2. RAIFFA H (1968) *Decision Analysis.* Addison–Wesley.
3. SPETZLER CS & STAËL VON HOLSTEIN C–AS (1975) Probability encoding in decision analysis. *Mgmt Sci.* **22**(3), 340–358.
4. STAËL VON HOLSTEIN C–AS (1973) Probability encoding in practice. In Proc. Conf. on *The role and effectiveness of theories of decision in practice* (eds WHITE DS & BOWEN KC) 148–157. Hodder and Stoughton.

SH Arcade

Department of Management Science
Imperial College
London SW7 2BX
UK

(*April* 1979)

On Normative Policy Models in Developing Countries

DEVELOPING countries pose a peculiar challenge to management science. The wide need-availability gaps, inefficient administrative systems and processes, poor performance of programmes—all indicate a tremendous scope for management science applications. The scope for management science intervention is the greatest in public systems, especially those focusing on basic human needs. The political-administrative culture, value orientations of individuals and inadequate infrastructure of these public systems, however, militate against the use of formal management science methods.

In this context, the search for appropriate management science style for developing countries assumes significance [2]. There is well-founded skepticism about the applicability of formal, normative decision models to public systems in developing countries [5]. One school of thought even maintains that adaptive and heuristic administrative decision processes are not only more appropriate but also more optimal than formal and normative approaches [1, 4].

Some observations of policymakers trying out a normative policy model indicate that such models are useful if employed in selective situations. The experience with one such model is reviewed here.

THE PROBLEM AND THE MODEL

The model was developed as an analytical policy formulation approach for essential commodities in India [3]. The term 'essential commodity' refers to mass consumption items which are frequently purchased and account for a substantial portion of the consumption expenditure of poor people. Government policymakers had resorted to several measures over a period of three decades to influence the availability, quality and prices of such mass consumption items.

The proposed normative model [3] was a structured and analytical approach which required the policymakers to go through the following steps:

(1) Selection of commodities requiring policy attention.

(2) Analysis of consumption, distribution, production and physical characteristics of these commodities.

(3) Specification of a range of policy objectives and possible policy instruments.

(4) Establishing the importance of various policy objectives for the selected commodities.

(5) Assessing the effectiveness of policy instruments in achieving various policy objectives.

(6) Derivation of the mix of potentially effective instruments for the selected commodities.

The procedure required systematic application of judgement and structured proforma were provided to aid the policymaker. The approach was favourably received when presented to policymakers in a seminar.

TRIAL AND OUTCOME

To assess the utility of the approach to policymakers in the field, four senior administrators were personally approached to 'try out' the approach by simulating the policy making for essential commodities selected by them. Two of the administrators were willing to discuss policy making in general but politely refused to simulate with the normative model. They did not have much faith in normative policy models. The other two administrators were willing to try out the normative model. One used it for arriving at policy regarding wheat and lentils; the other for devising policy regarding edible oil and soap. They found the normative approach easy to use, logical and stimulating. They made several suggestions for the improvement and application of the normative model:

(1) Background material in the form of empirical data about the commodities and past effectiveness of various policy instruments should be provided to policymakers.

(2) A glossary of terms used in the normative model should be provided so that these terms are uniformly understood and applied.

(3) The model is likely to be more useful in situations where previous policy experiences are limited.

(4) The model can be useful at various levels of policy making and implementation.

(5) The systematic normative approach can be a valuable training device for inducting new officials in the machinery of policy making and implementation.

CONCLUSION

To the extent that the above observations are generalizable, it appears that in developing countries, normative policy models are useful to some policymakers for selective purposes, if empirical and expository backup is provided. If this be the case, the task of the management scientist in developing countries is to discover 'zones of vulnerability' and inject management science inputs through those zones.

REFERENCES

1. GORE WJ (1964) *Administrative Decision Making: A Heuristic Model.* John Wiley, New York, USA.
2. HINDLE D (1978) Management science in developing countries: adopting the right style. *Omega* 6(4), 370.

3. KHURANA R, BHANDARI L, VORA MN & DHOLAKIA N (1979) An analytical approach for formulating distribution policy. In *Public Distribution System: Evolution, Evaluation and Prospects* (eds DHOLAKIA N & KHURANA R) Oxford & IBH, New Delhi, India.
4. LINDBLOOM CE (1965) *The Intelligence of Democracy: Decision Process Through Adjustment.* Free Press, New York, USA.
5. SIFFIN WJ (1976) Two decades of public administration in developing countries. *Pub. Admin. Rev.* **36**(1), 61–71.

N Dholakia
R Khurana
L Bhandari
MN Vora

Indian Institute of Management
Vastrapur
Ahmedabad 380015
India

(*March* 1979)

Vol. 7
1979

Normal Approximations for the Greater of Two Normal Random Variables

IN PRODUCTION or project applications of network analysis, one often encounters situations such that the completion time t is stochastic and knowledge of the probability density function (*pdf*) of t is required. In networks consisting of parallel paths, t is determined by the maximum of the time required for events on each path.

We will further define the setting by requiring that t be determined by the completion of independent processes. Thus the time to completion is the greater of the times to completion of the two processes, t_a and t_b. The determination of the *pdf* of t then depends upon the probability distributions for the timing of these two processes. We will assume their distributions are normal and will specify conditions under which the *pdf* of t is approximately normal.

Deriving the true distribution

Setting

$$t = \max(t_a, t_b),$$

let us identify the cumulative distribution functions of t_a, t_b and t as F, G and H respectively. Finally, let:

$$f(t) = F'(t),$$

From order statistics, we know that

$$H(t) = F(t)G(t),$$

and hence

$$h(t) = F(t)g(t) + f(t)G(t).$$

If both $f(t)$ and $g(t)$ are normal distributions, $F(t)$ and $G(t)$ are integrals which cannot be expressed in closed form. The form of $h(t)$ is, therefore, complex and difficult to manipulate.

The normal as an approximation

Clark [1] has provided formulas for determining moments of the distribution of t. However, as explained earlier, applications often require complete specification of the *pdf* and there is therefore a need to determine a manageable approximation to $h(t)$.

The existence of a close and easily developed approximation depends upon the relative locations and relative dispersions of the distributions of t_a and t_b. We seek combinations of means and standard deviations of the distributions of t_a and t_b such that we can approximate the distribution of t by a normal distribution with mean and standard deviation as found with Clark's formulas.

Because relative means and standard deviations are the determining factors, we can choose one of the two normal variables to have mean zero and standard deviation one. Where $E(t_a) < E(t_b)$, we arbitrarily identify t_b as the standard normal variable and allow the parameters of the normal distribution of t_a to vary. Each pair of values for the mean and standard deviation for t_a provides a case for testing the appropriateness of the approximation.

To begin, we used Clark's formulas to calculate the mean and standard deviation of the distribution of t with 176 different combinations of the mean and standard deviation of t_a. The 176 situations represent combinations of 16 means and 11 standard deviations for t_a. The values calculated are presented in Table 1.

The next task was to examine the prospects for using a normal distribution with the parameters from Clark's formulas as a surrogate for $h(t)$. The general approach was within the context of a goodness-of-fit test. However, it should be made clear that a statistical test in the ordinary sense is inappropriate. We *know* that $h(t)$ is *not* normal. Rather, we interpret a low level of significance for rejecting normality as implying that the *degree* of difference between $h(t)$ and the normal against which it is compared is relatively large and vice versa.

We took each 'hypothesized' normal and cut it into bands or histograms with width of one-tenth of a standard deviation. Then we calculated the area in each of these bands. The next step was to use $H(t)$ to determine the actual area of bands with the same locations as the ones in the 'hypothesized' normal. If the 'hypothesized' normal were identical to $h(t)$, the corresponding band areas from the two distributions would be identical.

The principal source of differences between the two distributions is the skewness of $h(t)$. For relatively large values of $\sigma(t_a)$ the source of the skewness is the density which is being 'thrown' to the right side of the distribution by dispersion. As $\sigma(t_a)$ grows smaller this type of skewness diminishes and the normal becomes a better approximation. However, when $\sigma(t_a)$ goes still smaller a new source of skewness appears. The distribution is becoming truncated, as it were.

The final step was to interpret the differences via a goodness-of-fit test. The results of the test can give a quantifiable indication of the relative accuracy of the approximation. Note that the difference figures are the expected values of the differences which would result from taking very large samples from the two distributions. We can, therefore, achieve internal consistency in the analysis by assuming the differences resulted from some arbitrary sample size; we chose $N = 10{,}000$. The test we used was the Kolmogorov–Smirnov one-sample test (see Conover [2]). Table 2 presents the results (note: H_0 is normality).

Clearly, for near-zero values of $E(t_a)$ and for large or very small values of $\sigma(t_a)$, the normal is a poor approximation. It is also interesting to note that the pattern mentioned above is very evident in Table 2. The rejections in the cells of the lower left-hand corner indicate the rejection pattern as $E(t_a)$ moves toward zero. The block of cells in the lower right-hand side indicate the rejection pattern as $\sigma(t_a)$ approaches zero.

An illustration

To illustrate practical use of Table 2 consider the following situation. The analyst determines two paths through a network that can be described as independent normal distributions. The time to completion is therefore the maximum of the two. Path A has a mean of 12 days and a standard deviation of 2.4 days. Path B has a mean of 16 days and a standard of 2 days.

TABLE 1. MEAN AND STANDARD DEVIATION OF $h(t)$ FOR $E(t_b) = 0$ AND $\sigma(t_b) = 1$)[1]

$E(t_a)$	$\sigma(t_a) =$ 2.0	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2	0.0
−3.0	0.09312	0.06652	0.04490	0.02826	0.01637	0.00862	0.00414	0.00186	0.00086	0.00048	0.00038
	0.99073	0.97906	0.97495	0.97642	0.98117	0.98694	0.99203	0.99558	0.99758	0.99850	0.99875
−2.8	0.11260	0.08244	0.05734	0.03747	0.02272	0.01267	0.00650	0.00316	0.00158	0.00093	0.00076
	0.99388	0.97833	0.97141	0.97130	0.97567	0.98202	0.98821	0.99293	0.99582	0.99724	0.99765
−2.6	0.13534	0.10144	0.07261	0.04915	0.03113	0.01831	0.01002	0.00524	0.00281	0.00175	0.00146
	0.99860	0.97851	0.96807	0.96569	0.96913	0.97573	0.98297	0.98899	0.99301	0.99512	0.99576
−2.4	0.16171	0.12393	0.09114	0.06379	0.04211	0.02605	0.01514	0.00847	0.00485	0.00319	0.00272
	1.00521	0.97994	0.96518	0.95974	0.96153	0.96789	0.97595	0.98333	0.98867	0.99167	0.99261
−2.2	0.19209	0.15035	0.11345	0.08194	0.05623	0.03647	0.02242	0.01335	0.00815	0.00562	0.00489
	1.01403	0.98296	0.96310	0.95368	0.95296	0.95836	0.96682	0.97545	0.98221	0.98627	0.98758
−2.0	0.22687	0.18115	0.14004	0.10419	0.07414	0.05025	0.03256	0.02056	0.01330	0.00960	0.00849
	1.02539	0.98796	0.96220	0.94783	0.94357	0.94708	0.95526	0.96479	0.97294	0.97815	0.97990
−1.8	0.26644	0.21678	0.17146	0.13116	0.09655	0.06820	0.04639	0.03093	0.02112	0.01589	0.01428
	1.03955	0.99533	0.96293	0.94262	0.93365	0.93412	0.94105	0.95084	0.96009	0.96640	0.96859
−1.6	0.31116	0.25770	0.20824	0.16350	0.12423	0.09117	0.06487	0.04547	0.03262	0.02549	0.02324
	1.05675	1.00542	0.96575	0.93852	0.92361	0.91968	0.92409	0.93315	0.94289	0.95003	0.95261
−1.4	0.36141	0.30435	0.25093	0.20185	0.15796	0.12010	0.08906	0.06536	0.04905	0.03969	0.03667
	1.07716	1.01854	0.97111	0.93608	0.91397	0.90412	0.90444	0.91139	0.92065	0.92809	0.93089
−1.2	0.41752	0.35715	0.30003	0.24687	0.19852	0.15594	0.12011	0.09192	0.07186	0.06000	0.05610
	1.10087	1.03495	0.97943	0.93587	0.90537	0.88801	0.88242	0.88552	0.89287	0.89974	0.90247
−1.0	0.47981	0.41648	0.35604	0.29915	0.24668	0.19964	0.15920	0.12653	0.10264	0.08815	0.08332
	1.12785	1.05477	0.99105	0.93845	0.89851	0.87207	0.85859	0.85574	0.85936	0.86441	0.86665
−0.8	0.54855	0.48270	0.41939	0.35926	0.30315	0.25213	0.20746	0.17061	0.14304	0.12598	0.12021
	1.15802	1.07805	1.00619	0.94428	0.89412	0.85716	0.83379	0.82264	0.82031	0.82196	0.82312
−0.6	0.62399	0.55610	0.49046	0.42769	0.36858	0.31422	0.26597	0.22550	0.19468	0.17530	0.16867
	1.19118	1.10469	1.02496	0.95371	0.89286	0.84424	0.80907	0.78718	0.77638	0.77270	0.77207
−0.4	0.70630	0.63692	0.56958	0.50483	0.44349	0.38661	0.33562	0.29235	0.25897	0.23774	0.23044
	1.22706	1.13448	1.04728	0.96694	0.89524	0.83420	0.78566	0.75066	0.72870	0.71752	0.71424
−0.2	0.79563	0.72534	0.65695	0.59100	0.52827	0.46982	0.41711	0.37207	0.33706	0.31464	0.30690
	1.26528	1.16711	1.07293	0.98392	0.90159	0.82783	0.76481	0.71465	0.67886	0.65780	0.65092
0.0	0.89206	0.82147	0.75272	0.68637	0.62317	0.56419	0.51090	0.46524	0.42967	0.40684	0.39894
	1.30546	1.20216	1.10155	1.00444	0.91196	0.82565	0.74765	0.68085	0.62879	0.59538	0.58382

TABLE 2. LEVELS OF SIGNIFICANCE FOR REJECTING NORMALITY

$E(t_a)$	$\sigma(t_a)$ 2.0	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2	0.0
−3.0											
−2.8											
−2.6											
−2.4	A										
−2.2	C										A
−2.0	E						B	B			E
−1.8	E					A	D	D	D	D	E
−1.6	E	B				B	E	E	E	E	E
−1.4	E	D				C	E	E	E	E	E
−1.2	E	E	A			C	E	E	E	E	E
−1.0	E	E	C			C	E	E	E	E	E
−0.8	E	E	D			A	E	E	E	E	E
−0.6	E	E	E	B		A	E	E	E	E	E
−0.4	E	E	E	D			E	E	E	E	E
−0.2	E	E	E	E	A		E	E	E	E	E
0.0	E	E	E	E	C		D	E	E	E	E

A means rejection is significant at 0.20 level
B means rejection is significant at 0.15 level
C means rejection is significant at 0.10 level
D means rejection is significant at 0.05 level
E means rejection is significant at 0.01 level

To begin we standardize the problem. Path *B* becomes the standardized normal. Path *A* becomes

$$E(t_a) = -2, \quad \sigma(t_a) = 1.2$$

Since that point lies in the 'accept' region the analyst can refer to Table 1 for appropriate parameters for $h(t)$, i.e.:

$$E(t) = 0.0741, \quad \sigma(t) = 0.9436.$$

We transform to return to the original scale and find the time to completion distribution can be approximated by a normal with a mean of 16.1482 days and a standard deviation of 1.8872 days.

REFERENCES

1. CLARK CE (1961) The greatest of a finite set of random variables. *Ops Res.* 145–162.
2. CONOVER WJ (1971) *Practical Nonparametric Statistics.* John Wiley, New York, USA.

Willis R Greer Jr
Gerald J LaCava

(*January* 1979)
The Amos Tuck School of Business Administration
Dartmouth College
Hanover
New Hampshire 03755
USA

Public Risk From Potential LNG Spills

ONE OF THE major issues in the siting of liquified natural gas (LNG) facilities is the possibility of accidental LNG spills, which may result in fatalities to the public. Clearly the likelihood of such spills and information about their potential consequences in terms of fatalities should be carefully examined before the facilities are licensed[1]. It is obvious that less populated areas create less public risk than more populated areas, given that other factors (e.g. probability of an accident) are equal. However, many people feel that an area with twice the population of another area presents twice the public risk. The purpose of this memorandum is to demonstrate that such a contention is false. In fact, in some circumstances, the population of one area can be 10 times another, and yet the public risk in terms of fatalities in the former case is only marginally greater than in the latter. The circumstances under which such implications are true are indicated in this note.

In order for the public to have any risk, a spill of LNG must occur. This may result, for example, from the collision of an LNG-carrier and another ship or from a land storage tank rupture due to an earthquake. Upon spilling, the very cold (−260°F) LNG begins to vaporize immediately. The resulting vapor in the correct mixture with air (5–15%) may be ignited. If an ignition occurs almost simultaneously with the spillage, a pool fire results until it is either extinguished or burns out. If immediate ignition does not occur, a vapor cloud, which is negatively buoyant, travels downwind a certain distance which depends on the spill size, its location, and atmospheric conditions. Until the vapor cloud becomes diluted to the point where it is no longer flammable (less than 5% natural gas to air), it may be ignited by several sources such as an open flame, a spark, or a very hot metal. If a cloud is ignited, the fire rapidly burns through it. Individuals within a vapor cloud at the time of ignition are in serious danger.

If an LNG facility is located such that no member of the public is exposed to the consequences of any LNG spill, then there is no public risk. However, the likelihood of finding such areas in the US, which also meet environmental standards, is not great. It is much more likely the case that some people will live within the vicinity of the proposed facility and/or shipping route. It is true for potential facility sites that a smaller population in an area implies a smaller public risk given other factors (e.g. weather) are equivalent. Hence, relatively unpopulated sites are preferred to more populated sites. However, the following somewhat counter initutive result below is also true.

Given that a vapor cloud of natural gas is traveling toward a populated area, an area with twice the number of people is *not* twice as endangered from the public risk point of view. It is always less than twice as endangered and in many situations, only marginally more so than the original area.

This result can be illustrated with a simple model of the vapor cloud ignition process. The model is in no sense a complete description of the ignition process. However, it does illustrate the aspect of the actual process which causes the result stated above. The qualitative insight gained from this is useful in identifying good facility sites. After presenting the model, its limitations and interpretation will be discussed.

The model assumes a vapor cloud is traveling toward a number of houses. It will, if not ignited, pass over these houses one at a time. Given the vapor cloud reaches a house, the probability that it is ignited by a source within that house is P. There are a total of N houses. If the vapor cloud is not ignited by one of these houses, it eventually disperses. The model assumes that once a vapor cloud reaches a house, that house will be in the cloud at the time it is ignited if an ignition occurs. All houses within a cloud at ignition are assumed destroyed.

With this model the cloud reaches the first house and is ignited with probability P. If it is not ignited, it travels to the second house where it is ignited with probability P. Thus the probability that it is ignited by the second house is $(1\text{-}P)P$, which is the probability $(1\text{-}P)$ that it is not ignited by the first house, multiplied by the probability P that it is ignited by the second house. Similarly, the probability that the vapor cloud is ignited by the n^{th} house it reaches is $(1\text{-}P)^{n-1}P$. The expected number of houses destroyed $\bar{n}(N)$ as a function of the number of houses N is calculated from

$$\bar{n}(N) = \sum_{n=1}^{N} P(1 - P)^{n-1} n, \tag{1}$$

which is the sum of the probabilities of the n^{th} house igniting the vapor cloud multiplied by the number of houses destroyed, n. If N is infinite, it follows from (1) that

$$\bar{n}(\infty) = \frac{1}{P}. \tag{2}$$

Thus regardless of how many houses are in the path of a vapor cloud, the expected number destroyed is less than $1/P$. Qualitatively, the reasoning for this is that the vapor cloud will very likely be ignited by the first several houses so it will not reach the further houses.

For finite N, (1) is expanded to give

$$\bar{n}(N) = \frac{1}{P} - (1 - P)^N \left[\frac{1}{P} + N\right]. \tag{3}$$

Substituting 100 and 1000 for N in (3) is insightful. If the probability of a vapor cloud being ignited by a source in any house is 0.1, reducing the number of houses in the path of a vapor cloud from 1000 houses to 100 only reduces the expected number of houses destroyed by approximately 0.03%. The expected number destroyed with only 20 houses in the path of a vapor cloud is approximately 65% of the expected number with 1000 houses in this path.

[1] Interested readers are referred to Keeney *et al.* [1] which describes in considerable detail the methodology employed in an actual study.

Vol. 7
1979

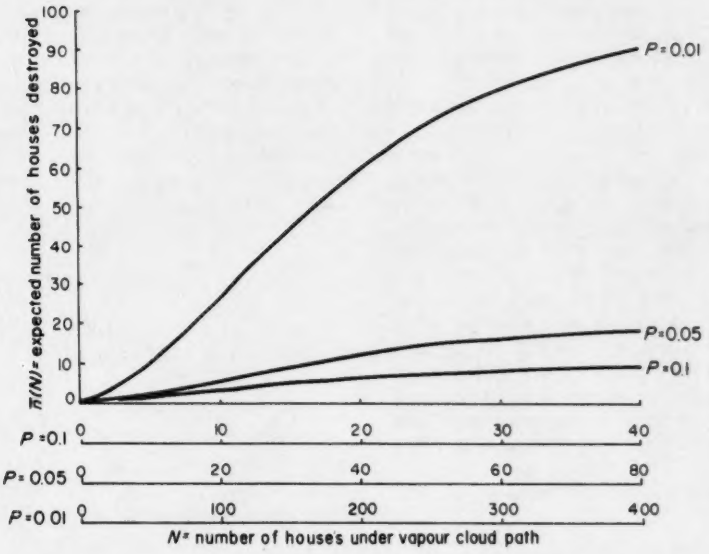


FIG. 1. *Expected number of houses destroyed given ignition probability P and number of houses N*

A plot of $\bar{n}(N)$ versus N for three different values of P is shown in Fig. 1. As can be seen, by the time that there are at least $4(1/P)$ houses in the path of a cloud, the additional potential damage due to more houses is limited to less than 10% of the total expected damage. To examine this, the number of houses, referred to as n', necessary to account for 50, 75, and 90 percent of the potential damage with an infinite number of houses is tabulated for different values of P. To do this, we calculate n' from

$$(1 - P)^{n'}\left[\frac{1}{P} + n'\right] = \frac{y}{100}\left[\frac{1}{P}\right],$$

A plot of n' versus P for $P \leq 0.2$ is given in Fig. 2. Thus, if $P = 0.05$, the presence of only 75 houses has a damage potential which is 90% as great as if there were 5000 (or an infinite number of) houses. The implications of wind direction can also be examined using Fig. 2. For example, if the probability of a house igniting a vapor cloud is $P = 0.05$, then 50 houses downwind a potential spill site 20% of the time is more dangerous to the public than 1000 houses downwind from the same site 10% of the time.

All of the quantitative statements made in this paper have been based on a simplistic model. For instance, the model assumes that the vapor cloud will pass over one house at a time. In reality, it would be the case that the cloud could pass over more than one house simultaneously. The model assumes that if a source of ignition in a home will ignite a cloud, it will occur immediately as the cloud reaches the home. In fact,

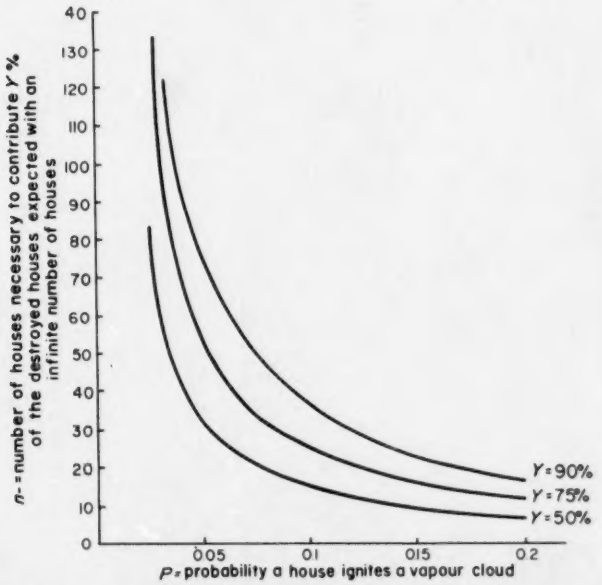

n' = number of houses necessary to contribute Y % of the destroyed houses expected with an infinite number of houses
Y = 90%
Y = 75%
Y = 50%
P = probability a house ignites a vapour cloud

FIG. 2. *The contribution to expected destroyed houses due to the first n' houses*

there would likely be a delay as the natural gas filtered into the home. The model also does not account for houses destroyed which may have been near, but not within, a cloud at the time of ignition.

Because of these limitations, the question is whether the model has any relevance to LNG risks. I feel the answer is definitely yes. Simplified modelling of complex processes plays an important role in analyses of public risk. One should not put great faith in the exact numbers resulting from the model. However, by varying input parameters (i.e. number of houses, probability of ignition) as done to yield the figures, crucial insights into the complex process can be obtained. In this case, a mathematically sophisticated model of the vapor cloud ignition process would yield results qualitatively similar to those presented here. In fact, many of the relative quantitative statements, e.g. the expected number of destroyed houses in one situation is twice that of another situation, are likely reasonable representations of reality. The use of the results should be to provide insight for selecting liquified natural gas facilities.

Clearly, if a completely unpopulated area is available for an LNG facility, it would be very safe from a public risk viewpoint. However, once the areas being considered for potential sites are even sparsely populated, the overall public risk of fatalitites in most situations will be controlled by the likelihood of accidential LNG spills and the weather in the vicinity more than the population. It is important to recognize this in screening to identify potentially good sites.

REFERENCES

1. KEENEY RL, KULKARNI RB & NAIR K (1979) A risk analysis of an LNG terminal. *Omega*, **7**(3), 191–205.

Ralph L Keeney

Woodward-Clyde Consultants
3 Embarcadero Center, Suite 700
San Francisco, CA 94111, USA

(*January* 1979)

A Goal Programming Model for Improved Transportation Problem Solutions

INTRODUCTION

GOAL PROGRAMMING has been successfully applied in many areas of management as an aid to decision making. Applications in accounting [2], finance [1, 4], marketing [3], and the management sciences [5] attest to its effectiveness in dealing with optimization problems. Yet, goal programming is still in its early age of development, and virgin areas of application are still being discovered. The area of transportation management discussed in this paper appears to be one of these virgin areas. The simplicity of operations research techniques, such as the transportation method, to deal with problems in this area has undoubtedly curtailed the entrance of more sophisticated techniques, such as goal programming. As the transportation problems become more complex, so must the problem-solving theory provided by the management sciences evolve to meet this challenge.

The purpose of this paper is to demonstrate the use of goal programming as a means to improved transportation problem decision making. Specifically, this paper has two objectives: (1) to show how a goal programming process is utilized in combination with the transportation method to improve decision making; and (2) to demonstrate how goal programming can improve cost minimization by utilizing additional information.

BACKGROUND OF THE PROBLEM SITUATION

A meat processing organization located in the Midwest controls a chain of restaurants operating over a five-state area. Hamburger products are the major food item served at these restaurants. As such, the organization seeks to minimize their costs of this product by utilizing the transportation method in shipping this product from processing plants to storage facilities.

The processing organization plans its production and shipping schedule on a monthly basis to take into consideration changes in demand and cost factors. The organization operates three processing plants and six storage facilities. The processing plant generates the raw hamburger and transports it in ten-pound packages to the storage facilities. The minimum number of units to be shipped from a plant to a storage facility is 400. If less than 400 are shipped, additional costs per unit might be incurred. At the storage facilities, the ten-pound packages are broken down into the various hamburger products and stored to meet the daily demand of the nearby restaurants. Since the processing cost per pound varies slightly between the processing plants, the product's processing cost is included with the transportation costs when analyzing cost information. The film gathers cost, supply and demand information on a monthly basis and determines, using the transportation method, the cost minimizing sources and destinations. Unfortunately, the transportation method is not structured to take into consideration certain types of additional information relevant to transportation management; for example, interval rather than static demand estimates. Supply and demand information utilized in the traditional transportation algorithm must be a static estimation, while supply information can easily be obtained, as it is usually, under the control of management. Static estimates of demand information are usually determined by first obtaining interval estimates from marketing managers and converting them to a representative midpoint or average value for the entire interval. If the interval estimate could be included in the analytical process, a further cost reduction might be incurred. To include the interval demand estimate information into the transportion problem, goal programming analysis was suggested.

The cost information used in the following goal programming analysis was obtained from the files of the meat processing organization. In examining the previous 12 months' applications of the transportation method, it was found that 75% (i.e. 9 out of 12) of the applications had the same cost per unit and per cell; table 1 displays the most typical situation facing the firm.

Goal programming model

In developing a model to improve the transportation method's solution several considerations must be made. First, the total supply of units is fixed and the model must constrain the solution to reflect this. Second, the supply of units per supply location or processing plants is fixed, and the solution should reflect this. At the same time as these allocations are being made, the model should force the selection of the set of allocations that best minimized cost. A third consideration in order to improve upon the transportation method's solution, is a weighting process that will force either increased or decreased allocations of units toward selective demand destinations that will reduce the average cost per unit. Determining these selective demand destinations involves several steps of a relatively simple comparative process. The steps are:

1 Compute the weighted average cost per unit according to the transportation method solution;

2 Compute the average cost per unit for each demand destination;

3 Determine the range between the highest and lowest value in step 2;

4 Divide the range by the number of demand destinations in the problem;

5 Add and subtract the value in step 4 to the value in step 1 to determine a cost interval;

6 Those values that fall below this interval represent demand destinations that should have increased allocations, because they will on the average decrease per unit costs; those values that fall above this interval represent demand destinations that should have decreased allocations, because they will on the average increase per unit costs.

The interval values are used instead of the simple average, since it will reduce problems connected with a total reallocation of resources. The intent of the model posed by this paper is to improve on the transportation model's initial solution, not create a new allocation. If the average value is used as the criterion, it increases the possibility of a complete reallocation by allowing all demand destinations to change their value.

Fourth, remaining demand destinations (i.e. those demand destinations whose cost is within the interval) should be forced to utilize the midpoint estimate as a demand figure in the final estimate. To meet the requirements posed above, a goal programming model is presented in Fig. 1.

FIG. 1. *A generalized goal programming model for a transportation problem*

$$\text{Minimize: } Z = P_1(d_1^- + d_1^+) + P_2 \sum_{i=1}^{m} (d_{i+1}^- + d_{i+1}^+) + P_2(d_{m+2}^+) + P_3(d_{j+m+2}^- \text{ where } j = 1, 2, \ldots, n \text{ for each positive } A) + P_3(d_{j+m+2}^+ \text{ where } j = 1, 2, \ldots, n \text{ for each negative } B) + P_4(d_{j+m+2}^- + d_{j+m+2}^+ \text{ where } j = 1, 2, \ldots, n \text{ for each negative } A \text{ and positive } B)$$

$$\text{Subject to: } \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} + d_1^- - d_1^+ = b^{TU}$$

$$\sum_{j=1}^{n} x_{ij} + d_{i+1}^- - d_{i+1}^+ = b_i \qquad (\text{for } i = 1, 2, \ldots, m)$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} + d_{m+2}^- - d_{m+2}^+ = b^{TC}$$

$$\sum_{i=1}^{m} x_{ij} + d_{j+m+2}^- - d_{j+m+2}^+ = b^{TD} \qquad (\text{for } j = 1, 2, \ldots, n)$$

$$\check{b}_j \leqslant \sum_{i=1}^{m} x_{ij} \leqslant \hat{b}_j \qquad (\text{for } j = 1, 2, \ldots, n)$$

and x_{ij}, all d^-, d^+'s $\geqslant 0$ (for $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$)

where:

x_{ij} = number of units to be transported from an ith supply location to jth demand destination
m = number of supply locations
n = number of demand destinations
c_{ij} = cost of a unit transported from an ith supply location to jth demand destination
A = an average demand cost per destination below the lower bound in the cost interval found by the difference between:

$$\left[\frac{\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}^*}{m \times n} - \frac{R}{n}\right] - \left[\sum_{j=1}^{n} \frac{\left(\sum_{i=1}^{m} c_{ij}\right)}{m}\right]$$

R = the absolute difference between the lowest and highest average demand cost per destination found by the absolute difference between:

$$\left| \left(\frac{\sum_{i=1}^{m} c_{ij}}{m}\right)_{\text{lowest}} - \left(\frac{\sum_{i=1}^{m} c_{ij}}{m}\right)_{\text{highest}} \right|$$

B = an average demand cost per destination above the higher bound in the cost interval found by the difference between:

$$\left[\frac{\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}^*}{m \times n} + \frac{R}{n}\right] - \left[\sum_{j=1}^{n} \frac{\left(\sum_{i=1}^{m} c_{ij}\right)}{m}\right]$$

P_i = goal programming priorities
d_i^-, d_1^+ = deviational variables
b^{TU} = total number of units transported from supply locations to demand destinations
b_i = number of units to be transported from each of the ith supply locations
b^{TC} = arbitrarily selected total transportation cost estimate
b^{TD} = estimated number of units demanded by each destination (based on midpoint or average *etc.*)
$\hat{b}$ = optimistic estimate of the number of units demanded by each destination
$\check{b}$ = pessimistic estimate of the number of units demanded by each destination

* *Denotes only* c_{ij} values that have cell allocations in accordance with the transportation method should be included in this summation.

ANALYSIS AND CONCLUSIONS

Using the data given in Table 1, the goal program was formulated and solved; the solution is presented in Table 2 along with solutions from the transportation algorithm and a straight linear programming formulation.

In comparing the transportation method with the goal programming solution, it can be seen that while both are feasible solutions (i.e. both have no allocation less than 400 units), the goal programming solution did reduce costs by an additional $110.00. This difference is rather insignificant relative to the total amount, but it should be kept in mind that the cost of the product is included in these figures. The result is to reduce the variability of cost between the possible allocations, since product cost does not vary significantly. In other words, the opportunity to improve cost significantly is not available because the problem lacks cost variability. In problems that have greater cost variability, greater improvement in cost reduction is possible. It should also be mentioned that while the linear programming solution results in the same minimized cost figure, the resulting allocation is infeasible, since one of the alloca-

Vol. 7
1979

TABLE 1. SUPPLY, DEMAND, AND COSTS INFORMATION

Processing plants	Storage facilities 1	2	3	4	5	6	Supply
1	4.3	4.0	4.4	4.7	4.9	5.2	3000
2	4.8	4.9	4.2	4.0	4.9	4.3	4800
3	4.9	4.3	4.6	4.1	4.1	4.8	7800
Demand	4160	3640	1560	2080	2600	1560	15,600
Interval low	4060	3540	1460	1980	2500	1460	
Estimates high	4260	3740	1660	2180	2700	1660	

tions is below 300 units. This result also indicates that, in this particular application, there is more than one optimal allocation, and if both the goal programming and linear programming problems had included constraints for the allocations for the decision variables to be greater than 400 units, they would have, in this case, arrived at the same solution.

It can be seen that these 6406 units are really the number of dollars over the $60,000 that the resulting solution requires. Priority 4's underachievement represents the failure of the objective function to force the remaining storage facilities (i.e. those with an average cost per unit within the cost interval) to use the midpoint estimate of 100 units each in storage facilities 2 and 6, and an increase in storage facility 3 of 100 units.

The use of this model in further cost reductions is quite informative in that managers know which storage facilities incur an increase in units and those which incur a decrease. With this information, storage facility managers might be able to revise their demand estimates to allow for further increases or decreases. Running the goal programming problem again with the revisions will invariably reduce per unit costs.

Another advantage to the goal programming approach is that it can easily include non-quantifiable information into the decision process. For example, suppose one of the storage facility's service area is experiencing a rapid growth rate in sales. The organization might prefer to have the storage facility receive the optimistic demand estimate. This can be accomplished by including that storage facility's d_i^- deviational variable in the objective function's third priority. The result may or may not increase cost. In the application of this paper the inclusion of this deviational variable would have increased costs if the increase is to occur in storage facilities 1, 2 and 6. Otherwise, the resulting cost would remain the same. At least with the goal programming solution, the decision makers would know immediately whether such a preference would cause an increase or not.

In conclusion, therefore, we have shown that it is possible to improve the transportation formulations by using goal programming and that this improvement has allowed the introduction of internal rather than merely point estimates.

Finally, the information provided by the goal programming formulation can be used to decide upon non-quantifiable preferences of the manager.

REFERENCES

1. KEOWN AJ (1978) A chance-constrained goal programming model for bank liquidity management. *Decis. Sci.* **9**, 93–106.
2. KORNBLUTH JSH (1974) Accounting in multiple objective linear programming. *Acctg Rev.* **49**, 284–295.

TABLE 2. SOLUTION COMPARISON BETWEEN THE TRANSPORTATION METHOD, GOAL PROGRAMMING AND LINEAR PROGRAMMING

Transportation in units from plant no. to storage no.	Transportation* method	Goal programming	Linear programming
x_{11}	3000	3000	3000
x_{23}	1560	1660	1460
x_{24}	1680	1680	1880
x_{26}	1560	1460	1460
x_{31}	1160	1060	1060
x_{32}	3640	3540	3740
x_{34}	400	500	300
x_{35}	2600	2700	2700
Total units shipped	15,600	15,600	15,600
Total cost ($)	66,516	66,406	66,406
Cost per unit	4.26	4.25	4.25

Turning to the goal programming solution itself, underachievement is incurred in Priorities 2 and 4, of 6,406 and 300 units respectively.

* Demand figures used were obtained from the midpoint of the intervals.

3. LEE SM & NICELY RE (1974) Goal programming for marketing decision: a case study. *Jl Mktg*, **38**, 24–32.
4. MORRIS R (1976) *Integer goal programming: methods, computations, applications.* Unpublished PhD Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.
5. SCHROEDER RG (1974) Resource planning in university management by goal programming. *Ops Res.* **22**, 700–710.

NK Kwak
MJ Schniederjans

(*December* 1978)

School of Business & Administration
Saint Louis University
2674 Lindell Boulevard
Saint Louis
Missouri 63108
USA

Vol. 7
1979

The Efficient Use of Non-Renewable Resources

THIS PAPER is about the efficient use over time of stocks of non-renewable resources. It begins by summarising the contemporary economic theories concerning the optimal rate of depletion, and it goes on to say why the methodology of these theories may be deficient. It is suggested that externalities, including political implications, need to be considered more explicitly in decision-making, and the paper goes on to illustrate this proposition by proposing a mechanism for the political distribution of North Sea oil revenues. The paper does not, for reasons of space, consider a number of issues which have traditionally been closely related to the question of optimal rate of depletion. These issues include questions of common ownership, the efficacy of the market mechanism, and problems of pollution.

Economists since at least the time of Adam Smith have regarded the development of economic systems over time as a race between technical progress on the one hand and limited natural resources on the other. But the importance which they have attached to this problem has fluctuated. Concern which existed amongst economists in 19th century Britain evaporated with the arrival from North America of supplies of low priced corn, made possible by a combination of improvements in the shipping technology and the cultivation of virgin lands. The issue has once again been brought to the forefront of public attention by recent much-publicised suggestions that the world as a whole may be running out of supplies of a particular class of natural resources, those which are exhaustible or non-renewable. While depletion of these stocks in a physical sense seems undeniable, most contemporary economists have been inclined to accept the fragmentary empirical evidence available which suggests that there has not been depletion in an economic sense. Technical change, economies of scale, and product-factor substitution have meant that extractive commodities have over the past 100 years or so become less scarce in terms of the sacrifices required to obtain them [8].

Nevertheless, a number of contemporary economists, [1,7,4] have in the last five years turned their attention to the question: at what rate should the non-renewable resources of a society be depleted? They have treated this question as an extension of a related question, already well established in the literature of economics: at what rate should the renewable resources of a society be accumulated, i.e. what is the optimal rate of investment over time? The answers traditionally given to these questions involve the maximisation of a sum of utility flows to be derived from future consumption. The principal parameters which occur in the models devised to answer these questions concern substitutability between factors of production, the rate and character of technical change, and uncertainty.

It is intuitively clear that the extent to which the non-renewable resource may be substituted by a renewable resource in the process of production is a critical element in determining the optimal rate of depletion. Even if the non-renewable resource is such that all production of final goods would fall to zero in its absence, the crux of the matter is the extent to which its use can be spread thinly over future time. We are operating here at the boundaries of production functions, where the economists' traditional assumptions of a constant elasticity of substitution between factors is not at all helpful. It simply begs the essential question. But there is seldom any empirical evidence to suggest what the substitution possibilities actually are. It is fair to say that those economists who have discussed this question appear to accept it as a working hypothesis that the elasticity of substitution is greater than or equal to 1.

Koopmans considers the extreme case of what he calls an essential resource, defined as one which is essential to sustaining life, is incapable of complete recycling and has no substitute either now or later within the remaining period of its availability. In this model the survival period becomes the policy variable, as does the population in some part of the future. However, Koopmans does not really believe an essential resource as strictly defined above really exists, though he recognises that it is a question on which scientists and engineers are more able to give an opinion than an economist.

A more plausible model is advanced by Dasgupta & Heal [1] who postulate a non-renewable resource for which a substitute may become available through discovery or invention, at an uncertain future date. They depart from conventional economic analysis in that they envisage technical progress not as a continuous but as a discrete event. But they assume that the arrival of the new technology is uninfluenced by policy, i.e. that the acquisition of knowledge is costless. The optimal path through time found by Dasgupta & Heal [1] consists of two successive segments. One is the path to be followed from $t = 0$ up to the as yet unknown time $t = T$ of the availability of a substitute. The other is the path to be followed from T onwards. Pindyck [9] has recently argued that the optimal rates of exploration and depletion of non-renewable resources are inter-related and must be jointly determined.

On uncertainty, given fairly typical assumptions, the optimal rate of extraction when the resource stock is uncertain is less than the optimal rate for the expected value of the stock. That is, uncertainty implies a more conservative extraction policy [2]. While economists have traditionally dealt with uncertainty by adding a margin to the rate at which future utility is discounted, Dasgupta & Heal [1] show that this procedure, when applied to the optimal depletion rate question, is only free of error in the special case where the substitute arrives at the moment when the stock of the exhaustible resource has no value, and when the technical change is sufficiently radical to cause the 'old' capital stock to lose all its value. The conclusion of Dasgupta and Heal is that even though resources may not remain non-renewable over the indefinite future, the substitutes may be a long time in coming, and the intervening generation may be that much worse off. This conclusion draws attention to the importance to the rate of discount and the notion of inter-generational equity. These are questions to which Koopmans and Solow [4] have drawn attention and their conclusions are

briefly as follows.

(1) The problem of whether and how much to discount future utilities cannot be equitably resolved *a priori* and in the abstract [4].

(2) The likelihood of the arrival of technological change should influence one's view of inter-generational equity [1]. (Dasgupta & Heal).

(3) Earlier generations are entitled to draw down the finite pool of non-renewable resources (optimally) so long as they add to the stock of reproducible capital (optimally) [13].

The idea that a substitute for a non-renewable resource which does not yet exist may arise some time in the future corresponds quite clearly to two significant economic activities: applied research leading to the development of a new technology (e.g. the breeder reactor) and exploration leading to the discovery of additional stocks of the resource. Contrary to the assumptions of many theoretical models of optimal depletion these activities are neither costless nor random. In the case of new technology, increased spending on applied research and development may be expected to bring nearer the time of its arrival, while in the case of exploration for non-renewable resources it can be argued that the size of the available reserves is a function of the price of the resource, for the simple reason that a rise in the resource price justifies recourse to deposits of a lesser quality and/or a greater inaccessibility.

We noted above the empirical findings which have supported the apparent complacency of economists in the face of the physical depletion of the world stocks of non-renewable resources. However, the interpretation of these findings needs to be qualified on two grounds. First, continuation of the trends observed in the recent past need not persist into the future. Secondly, the methodology of these studies ignores the environmental cost of resource extraction, transport, and processing. It is possible that these have risen sufficiently to offset the decline in market costs. There are other externalities which are likewise not considered and which may be of even greater importance.

One such externality which is overlooked by the conventional methodology is the political consequences of the expansion of the nuclear energy programme. It is well-known that the technological substitute for non-renewable energy sources is the production of atomic energy through nuclear fission. New risks are asociated with the introduction of this technology, viz. the possible failure of emergency cooling systems and the production of plutonium and long-lived radio active fission products. To prevent disasters arising from the activities of criminals, psychopaths, or fanatics "will entail an unprecedented extension of the internal and international security system", [6]. The foregoing is a specific example of the more general type of externality which Schumacher [11] associates with industrialisation. He asserts that there are three categories of non-renewable resource which the industrial system consumes. These are "fossil fuels, the tolerance margins of nature, and the human substance". The "tolerance margins of nature" refers to those substances compounded by scientists, which are 'unknown to nature', and against many of which 'nature is virtually defenceless'. The third category refers to something which 'perhaps cannot be measured at all' except for certain symptoms of loss, indicated by statistics of crime, drug addiction, vandalism, suicide etc., etc. Such externalities are generally ignored in contemporary economic theorising, perhaps because they cannot be quantified. As they do not lend themselves to quantification, the relationships are seldom specified. To put the matter another way: contemporary economic theory continues to assert, and it seems politically to be widely accepted, that the welfare of human societies is a monotonically increasing function of the volume of production of material goods—despite all the empirical evidence to the contrary. If we accept Schumacher's classification of "human substance" as part of the world's stock of non-renewable resources, then this proposition may be germane to the debate.

Turning from the broad question of the world's stock of non-renewable resources to the narrower issue of the stock of oil in the North Sea, we can see again how the conventional methodology of economic analysis fails to take account of factors which appear to be decisive in determining the rate of depletion in practice. A clear exposition of the parameters which should determine the optimal rate of depletion in the context of the North Sea is set out in the paper by Robinson & Morgan [10]. In the same paper however three different rates of depletion of the North Sea oil stock are revealed in practice. These are (a) the depletion rate in the UK sector prior to 1974, (b) the depletion rate in the UK sector since 1974, and (c) the depletion rate in the Norwegian sector. It is evident that the differences between these rates are not satisfactorily explained by the conventional theoretical model.

Since the costs and benefits of a particular rate of depletion policy are not uniformly distributed throughout the population surrounding the North Sea basin, nor uniformly through time, the question of the clients for the optimality criterion is of considerable importance. In other words, questions concerning the optimal rate of exploitation cannot be answered without first answering the question 'optimal for whom?' This tends to be glossed over in the literature, though Koopmans and Solow [4, 13] have considered it. In practice, the present rate of exploitation in the UK sector is hardly likely to be optimal for the inhabitants of Shetland. While official Norwegian policy may be optimal for the present generation of inhabitants of the Norwegian seaboard, only time will tell whether it is optimal for future generations of Norwegians, wherever they reside.

From the foregoing observations it should be clear that, in the writer's opinion, this subject area is ripe for development in the direction of political economy. Although the existing methodology can undoubtedly make a contribution, it is a limited one, since the most interesting questions of policy in this area do not lend themselves readily to quantitative analysis. Of course this statement is not peculiar to the issue of the depletion rate of non-renewable resources: much the same could be said about the question of the costs and benefits of UK entry to the common market. It would be wiser to recognise this explicitly, and try to make sure that the methodology of a 'new' political economy is an improvement upon the methodology of the old.

Perhaps what is meant by political economy in the context of non-renewable resources in the North Sea may be illustrated by a proposed mechanism for the political distribution of some of the benefits arising. The rest of the paper is devoted to outlining this proposal.

The discovery of oil in the Scottish waters of the North Sea which in political terms dates from around 1972 has had a remarkable effect on Scottish/English relations. While it has removed the principal constraint

Vol. 7
1979

on political independence it has created in England a hardening resistance to that idea, based upon the concern that loss of the North Sea oil reserves might put in doubt the capacity of the British government to refinance the heavy external borrowing of recent years. Thus the discovery of oil appears to have set the stage for a conflict of interests between the two countries where none before existed. It seems to be commonly supposed that the conflict of interests can be cast in the form of a zero sum game: whatever Scotland gains England must lose. All that can then be negotiated is the percentage share of a fixed sum. In fact some economic analysis can indicate that the conflict may not be as direct as is commonly supposed, to the extent that England's influence lies primarily in the direction of access to a large pool of foreign currency with which its international debts can be settled, while Scotland's interests lie in the acquisition of real resources which can be used to reconstruct its economy. A scheme can be devised to give England access to foreign exchange, while Scotland has access to the real resource content of the oil revenues. In practice, of course, England will want some of the real resources and Scotland will want some of the foreign exchange, but the presentation here is deliberately simplified in order to get across the main point, which is that there can exist some arrangement according to which the interests of the two countries may be complementary rather than conflicting. Nor is it intended that the scheme which is described below should be taken too literally: it is merely a sketch of the possibilities open to a wide range of modification in practice.

It should be made plain that the difference in size accounts for a great difference in the potential impact of the oil revenues on the two countries. In Scotland's case, the officially estimated tax revenues from oil will amount in the early 1980's to about 50% of her GDP while the same sum amounts to less than one third of the UK public sector borrowing requirement. Thus, while the oil revenues completely alter the prospects for rebuilding the Scottish economy they can only be of limited relevance in solving the economic problems of England. Let us suppose that the share of the foreign exchange earnings which accrue from the export of oil to a British government in the mid 1980's will be £3 billion per annum. Suppose further that a sovereign Scottish state comes into existence in the early 1980's and that the oil revenues accruing to this state are paid in dollars into an account jointly held by the Bank of England and the Central Bank of Scotland. Then the following arrangements might be agreed: that the Bank of England may draw on this account whenever it requires foreign exchange with which to repay its international obligations up to an agreed maximum limit. Whenever it makes a drawing on this account however it must transfer an equivalent amount in £s sterling at a rate of exchange to be agreed in advance to another account held by the Central Bank of Scotland. The Central Bank of Scotland would then be free to dispose of its sterling account together with its own portion of the joint foreign currency account according to the principles of economic policy laid down by the Scottish government. Of course, it would not be expected to convert these sterling holdings into dollars. To do this, would be to negate the purpose of the scheme. It should be noted that we are assuming a flexible rate of exchange between the Scottish pound and the pound sterling, and between the pound sterling and other major currencies. Provided that the pound sterling is not expected to fall significantly in value relative to other currencies, it should be possible to agree in advance a special fixed rate at which 'oil dollars' are to be converted into pounds sterling for the use of the Scottish Central Bank.

There are broadly three ways in which sterling funds accruing from the oil revenues to the Scottish government could be used:

(a) they could be devoted to purchasing material goods and services from current production in England.

(b) they could be devoted to the purchase of existing assets (land, property, factories) owned by sterling holders in Scotland and abroad.

(c) they could be devoted to loans from the Scottish to the English government.

The foreign exchange held in the joint account and not taken up by the Bank of England could be transferred to pounds sterling invested, lent abroad, given away as foreign aid or used for acquiring goods from overseas. The amount of the sterling funds which could be devoted to (a) would be limited by the capacity of the Scottish economy to absorb goods and services, and the capacity of the English economy to deliver then without undue strain. From the point of view of the English economy, it might be reasonable to expect it to be capable of exporting something of the order of £1 billion worth of additional goods and services to Scotland per annum. This would represent aroung 10% of its 1973 level of exports of goods and services. A 10% increase in exports would of course stimulate the level of aggregate demand in England but it would nonetheless represent a real burden on the economy of that value unless resources hitherto unemployed were brought into production. The acquisition of (b) existing assets, could begin in Scotland with the purchase of ownership of the large tracts of land that are in non-Scottish hands as well as the acquisition of minority shareholdings in productive enterprises operating in Scotland. There is the additional possibility of purchasing real assets outside Scotland. It is assumed of course that the sellers are willing to accept payment in sterling for their assets.

The remaining portion of the sterling account could in principle be devoted to (c) loans from the Scottish to the English government. The terms and conditions of these loans would of course be subject to negotiation and the demand would vary according to conditions in the English economy. What is left could be invested abroad or simply given away free as foreign aid. On the other hand, one could regard any unused funds as indicative of excess oil depletion. If the rate of depletion were reduced, then the oil would be left as a form of capital in the ground rather than transformed into a different capital stock.

REFERENCES

1. Dasgupta P & Heal GM (1974) The optimal depletion of exhaustible resources. *Rev. Econ. Stud.*
2. Gilbert RJ (1979) Optimal depletion of an uncertain stock. *Rev Econ. Stud.* **XLVI**(1), 142.
3. Harrigan FJ (1977) Economic Rent and Government Policy, 1964–1976, in the North Sea Province. MA dissertation, University of East Anglia, UK.
4. Koopmans TC (1976) Concepts of optimality and their uses. *Math. Progm*, **XI**(3).

5. MANNE AS (1974) Waiting for the breeder. *Rev. Econ. Stud.*
6. MISHAN EJ (1977) *The Economic Growth Debate.* London, UK.
7. PEARCE D (ed.) (1976) *The Economics of Natural Resource Depletion.* Macmillan, London, UK.
8. PETERSON & FISHER (1977) The exploitation of extractive resources. *Econ. Jl*, **87**(348).
9. PINDYCK RS (1978) The optimal exploration and production of non-renewable resources. *Jl Polit. Econ.* **86**(5).
10. ROBINSON C & MORGAN J (1977) *The Optimal Rate of Depletion of North Sea Oil.* Trade Policy Research Centre, London, UK.
11. SCHUMACHER E (1973) *Small is Beautiful.* Blond and Briggs, London, UK.
12. SIMPSON D (1976) Scotland, England and North Sea oil. In *The Radical Approach* (ed. KENNEDY G). Palingenesis Press, Edinburgh, Scotland, UK.
13. SOLOW RM (1974) Intergenerational equity and exhaustible resources. *Rev. Econ. Stud.*

DRF Simpson

(*April* 1979)

Director
The Fraser of Allander Institute
University of Strathclyde
100 Montrose Street
Glasgow G4 0LZ
Scotland, UK.

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 4, pp. 375 to 378
© Pergamon Press Ltd 1979. Printed in Great Britain
0306-0483/79/0801-0375$02.00/0

International Register of Theses and Dissertations

The following are obtainable at the Lyon Playfair Library, Imperial College of Science & Technology, Exhibition Road, London SW7 2BX, UK:

ABARA NC (1977) Risk and Uncertainty in Oil and Gas Projects in the North Sea. MSc Thesis, Imperial College, London.

AGRAWALA PK (1977) Analysis and Inflation Adjustment of Profitability. MSc Thesis, Imperial College, London.

ALCROFT TK (1977) Stock Control and the Cost of Shortages in the Post Office Telecommunications Business. MSc Thesis, Imperial College, London.

AVIS JM (1977) The Use of Control Theory to Analyse a Model of the London Metal Exchange Copper Market. MSc Thesis, Imperial College, London.

BAIRD AM (1977) A Project Management Game. MSc Thesis, Imperial College, London.

BASU RT (1977) Forecasting Leisure Markets. MSc Thesis, Imperial College, London.

BEANEY DG (1977) Container Retentions. MSc Thesis, Imperial College, London.

BELO-OSAGIE SU (1977) Quality Control Charts for Reliability of Large Computers. MSc Thesis, Imperial College, London.

BHATTACHARYYA S (1977) The Application of Minicomputers to the Management of Small Companies. MSc Thesis, Imperial College, London.

BROWN CE (1977) Management Problems of a Small Firm. MSc Thesis, Imperial College, London.

CHANG TCS (1977) Assembly Line Layout Problem. MSc Thesis, Imperial College, London.

CHAZAPIS L (1977) Minicomputers for Production Planning and Control. MSc Thesis, Imperial College, London.

CLARKE HB (1977) Business Location: Incentives and Constraints of South East England. MSc Thesis, Imperial College, London.

COLAH KH (1977) A Comparison of the Performance of British Firms and US Subsidiaries in the UK. MSc Thesis, Imperial College, London.

COX JRS (1977) An Evaluation of Retail Credit in Department Stores. MSc Thesis, Imperial College, London.

DEMADES DI (1977) Advancing Creativity in Industry. MSc Thesis, Imperial College, London.

DESPOTAKIS C (1977) The Chinese Postman's Problem and its Applications. MSc Thesis, Imperial College, London.

DOHA N (1977) Exchange Rate Fluctuations and the Management of International Operations. MSc Thesis, Imperial College, London.

FARR RW (1977) The Time-Constrained Travelling Salesman Problem: An Adapted Branch-and-Bound Algorithm. MSc Thesis, Imperial College, London.

FIRTH MH (1977) The Applications of Permanent Magnets to the Motor Industry. MSc Thesis, Imperial College, London.

GALVAO RD (1977) The Optimal Location of Facilities on a Network. PhD Thesis, Imperial College, London.

GILBERT SB (1977) The Paper Industry: A System Dynamics Model. MSc Thesis, Imperial College, London.

GLOSSOP MJR (1977) Marketing an Innovatory Design of Sports Rowing Boat. MSc Thesis, Imperial College, London.

GODOY T (1977) Managerial Aspects of the Consumers Co-Operatives in Great Britain. MSc Thesis, Imperial College, London.

GOYAL N (1977) Iran—Growth, Development and Trade, 1977–1983. MSc Thesis, Imperial College, London.

HAJIPAPAPHOTIOU P (1977) Some Aspects of the Capital Pricing Model. MSc Thesis, Imperial College, London.

HARIHARAN A (1977) The Term Structure of Commodity Prices: Some Applications of Stochastic Calculus. PhD Thesis, Imperial College, London.

HILL RM (1977) Forecasting of Bank Clearing Statistics. MSc Thesis, Imperial College, London.

HYND AJ (1977) Clustering Methods for Multiple-Depot Location. MSc Thesis, Imperial College, London.

INCE M (1977) Marketing Aspects of Overseas Contracts. MSc Thesis, Imperial College, London.

KANDIAH P (1977) Analysis of Pricing Policy: A Case Study. MSc Thesis, Imperial College, London.

KELLAS BB (1977) The Solution of a Depot Location Problem Using a Hybrid Tree Search/Dynamic Programming Algorithm. MSc Thesis, Imperial College, London.

KESHAVARZ E (1977) Transportation Methods and Routes to Iran: An Evaluation. MSc Thesis, Imperial College, London.

KWAN KC (1977) A Critical Examination of the Relationships Between Various Cyclical Indicators and the Usefulness of Leading Indicators in Forecasting. MSc Thesis, Imperial College, London.

LAM HM (1977) Approximations for the Tail of the Waiting Time Distribution in Single-Server Queues. MSc Thesis, Imperial College, London.

LANGDON JA (1977) A Graph Theoretic Model of Multinational Finance. MSc Thesis, Imperial College, London.

LEE JR (1977) The Electricity Supply Industry in England and Wales: A Survey. MSc Thesis, Imperial College, London.

LEE SH (1977) Weak Form Tests of the Efficient Capital Market Hypothesis. MSc Thesis, Imperial College, London.

LEE YC (1977) Incomes Policy and Pay Movements. MSc Thesis, Imperial College, London.

LEONG WT (1977) Uses and Consequences of Import Controls. MSc Thesis, Imperial College, London.

MCCARTHY WD (1977) Arbitrage and the Snake. MSc Thesis, Imperial College, London.

MERCER-DEADMAN MS (1977) Some Aspects of Management Information Systems and Data Collection. MSc Thesis, Imperial College, London.

NUNN RA (1977) Trading in Commodities. MSc Thesis, Imperial College, London.

PACKER LA (1977) Operating Costs of Vehicles in a Road Transport Fleet. MSc Thesis, Imperial College, London.

PAPALOIZOU PC (1977) The Economy of Cyprus. MSc Thesis, Imperial College, London.

POON LYM (1977) On the Solvency Margins of Insurance Companies. MSc Thesis, Imperial College, London.

QUINN JJ (1977) An Evaluation of the Methodologies Available for Forecasting the Demand for Washing Machines. MSc Thesis, Imperial College, London.

RICHARDSON MJ (1977) An Analysis of Passenger Car Tyre Demand in Western Europe up to 1985. MSc Thesis, Imperial College, London.

SACKRULE WN (1977) Current and Future Trends in the Biomedical Equipment Market in the EEC. MSc Thesis, Imperial College, London.

SCOTHERN JC (1977) The Changing Pattern of UK Trade and its Financing Demands and Methods. MSc Thesis, Imperial College, London.

SIMISTER CD (1977) British Banking and Financial Institutions: A Review. MSc Thesis, Imperial College, London.

SIMS CH (1977) A Review of the Methods Used in Staff Appraisal and Personnel Selection. MSc Thesis, Imperial College, London.

SPENCER MS (1977) Some Aspects of Pollution Control. MSc Thesis, Imperial College, London.

TAN SH (1977) Labour Mobility. MSc Thesis, Imperial College, London.

THOMAS DG (1977) Raw Materials for the Steel Industry. MSc Thesis, Imperial College, London.

THOO CF (1977) A Financial Model of a Leasing Company. MSc Thesis, Imperial College, London.

WEA CL (1977) A Study of UK Imports. MSc Thesis, Imperial College, London.

WELSH LJ (1977) The Market for Security Devices in the EEC Countries. MSc Thesis, Imperial College, London.

WHITLEY DA (1977) Feasibility Study of a Viewdata Based Information System for the Film and TV Industries. MSc Thesis, Imperial College, London.

WILDMAN MJ (1977) The Foreign Exchange Market: A Statistical Analysis of Exchange Rate Variability. MSc Thesis, Imperial College, London.

WISE RJ (1977) Some Aspects of Shiftworking. MSc Thesis, Imperial College, London.

The following are obtainable at the Interlibrary Loans Section, JB Priestly Library, The University of Bradford, Bradford 7, West Yorkshire, UK:

ARNOLD AJ (1976) Pergamon Press Limited and its implications for the accountancy profession. MSc Thesis, Management Centre, University of Bradford.

ASHTON DJ (1971) New Town Policy and the Employment of Manual Workers in the London New Towns, 1956–1964. MSc Thesis, Management Centre, University of Bradford.

BENSON ML (1972) Interpersonal Communications and Personal Influence in the Menswear Market: A Pilot Survey. MSc. Thesis, Management Centre, University of Bradford.

BLENKHORN DL (1977) The Status of the Materials Management Concept. PhD Thesis, Management Centre, University of Bradford.

BLOOMFIELD AM (1978) The Firm and Demographic Change. MBA Thesis, Management Centre, University of Bradford.

Vol. 7
1979

Bolade O (1976) Economic Planning in the Electricity Supply Industry. MSc Thesis, Management Centre, University of Bradford.

Chan JYH (1977) Expectancy theory as a Predictor of Work Behaviour: An Empirical Theoretical and Methodological Appraisal. MSc Thesis, Management Centre, University of Bradford.

Clegg SR (1974) Power in Organization Theory. PhD Thesis, Management Centre, University of Bradford.

Collins JB (1975) Management Labour Relations in the Scottish Woollen Industry, 1830–1970. PhD Thesis, Management Centre, University of Bradford.

Cometto G (1977) The Product Life Cycle: An Empirical Investigation. MBA Thesis, Management Centre, University of Bradford.

Cowell DW (1970) Automatic Vending as a Method of Distribution in the United Kingdom. MSc Thesis, Management Centre, University of Bradford.

Darwin DE (1974) A Study of Money and Job Enrichment as Motivation to Work. MSc Thesis, Management Centre, University of Bradford.

Davis D (1973) The Nature, Classification and Use of Some Methods of Mass Production. MSc Thesis, Management Centre, University of Bradford.

Deacon DJ (1978) The Educational Service and Local Authorities. MBA Thesis, Management Centre, University of Bradford.

Eldridge DA (1971) The Forecast of 1981 Car Ownership in London. MSc Thesis, Management Centre, University of Bradford.

Ellis SF (1975) Hypermarkets in France: Toulouse an Example. MSc Thesis, Management Centre, University of Bradford.

Finney MJ (1972) The World Petroleum Industry. MSc Thesis, Management Centre, University of Bradford.

Firth MA (1973) A Study of Capitalisation Issues in the UK during the three years 1968, 1969 and 1970. PhD Thesis, Management Centre, University of Bradford.

Giles JR (1977) Risk and Uncertainty in the Marketing of New Products. MBA Thesis, Management Centre, University of Bradford.

Gill HS (1967) Industrial Relations in the United Kingdom Atomic Energy Authority. PhD Thesis, Management Centre, University of Bradford.

Gold JG (1973) The Giveaway Industry. MSc Thesis, Management Centre, University of Bradford.

Goodman (1976) An Analysis of the Pattern of Multinational Joint Venture Behaviour, 1945 to 1975. MSc Thesis, Management Centre, University of Bradford.

Graves RH (1977) Higher Education for Management. MSc Thesis, Management Centre, University of Bradford.

Hall DS (1972) The Determination of Optimum Shop Size: A Study of Electrical Retailing. MSc Thesis, Management Centre, University of Bradford.

Halstead AS (1971) Exporting and the Multinational Company: A Study of the Institute of Export. MSc Thesis, Management Centre, University of Bradford.

Hawkins PJ (1976) Financial Fraud: A Consideration of its Control with examples involving British Companies. MSc Thesis, Management Centre, University of Bradford.

Herridge PA (1976) Planning in the Smaller Firm: a Case Study Analysis. MSc Thesis, Management Centre, University of Bradford.

Heyland WBW (1977) International Market Assessment and Selection. MBA Thesis, Management Centre, University of Bradford.

Hill MG (1977) The Training of Accountancy Students: An Investigation of Accountancy Education with Particular Emphasis on the Private Sector and its Present Role. MBA Thesis, Management Centre, University of Bradford.

Hillier TJ (1972) Decision Making in the Industrial Buying Process. PhD Thesis, Management Centre, University of Bradford.

Hillier TJ (1970) The Growth and Development of 'Out-of-Town' Shopping Centres in the United Kingdom: A Marketing Viewpoint. MSc Thesis, Management Centre, University of Bradford.

Hinchcliffe AW (1976) Development of a System for Meter Reading. MSc Thesis, Management Centre, University of Bradford.

Jackson J (1975) University Choice: A Model of Consumer Behaviour. MSc Thesis, Management Centre, University of Bradford.

Janssen MAC (1977) Government-Supported Export Financing and Export Credit Insurance Institutions in the UK and France: a Comparative Study. MBA Thesis, Management Centre, University of Bradford.

Marshall PA (1977) A Review of Linear Discriminant Analysis with Particular Reference to the Motor Oil Market in the United Kingdom. MSc Thesis, Management Centre, University of Bradford.

Mottram P (1977) Applications of Mathematical Programming to Agricultural Planning. MBA Thesis, Management Centre, University of Bradford.

O'Donnell J (1977) Aids for the Small Firm within the Construction Industry. MBA Thesis, Management Centre, University of Bradford.

PEACE L (1975) Flexible Working Hours. MSc Thesis, Management Centre, University of Bradford.

PERKINS DC (1977) A Study in the Burton Group Ltd. 1900–1977. MBA Thesis, Management Centre, University of Bradford.

PETTIFER DG (1977) Focus on Financial Fraud. MBA Thesis, Management Centre, University of Bradford.

PLATTS R (1973) The National Industrial Relations Court. MSc Thesis, Management Centre, University of Bradford.

PRICE DT (1976) Industrial Financing and Investment: a Comparative Analysis between the UK and West Germany. MBA Thesis, Management Centre, University of Bradford.

RADCLIFFE P (1977) Management and Education: Current Problems, Future Needs and Possibilities. MSc Thesis, Management Centre, University of Bradford.

RAISWELL JE (1978) Pricing Strategy and the Dynamics of a Shipbuilding Company. PhD Thesis, Management Centre, University of Bradford.

RYAN MC (1975) The Development of Supervisory Training: a Case Study—the Shipbuilding Industry. MSc Thesis, Management Centre, University of Bradford.

SALKILD MJ (1968) Buying Houses: their Operation and Opinions. MSc Thesis, Management Centre, University of Bradford.

SANDERSON P (1974) Capital Incentive Schemes. MSc Thesis, Management Centre, University of Bradford.

SANDERSON R (1977) Appraising the Use of Computer Assisted Management Games. MBA Thesis, Management Centre, University of Bradford.

SAVAGE GP (1975) Consumer Behaviour in the Market for Personal Bank Services. MSc Thesis, Management Centre, University of Bradford.

SCOTT P (1977) Job Enrichment, Job Design and the Two Factor Theory: a Re-evaluation. MSc Thesis, Management Centre, University of Bradford.

SILLS AF (1975) A New Approach to Homelessness: a Study of Management in Local Authority. MSc Thesis, Management Centre, University of Bradford.

STILL M (1971) Turnover of Asian Immigrant Labour in the Wool Textile Industry of the Bradford Region. MSc Thesis, Management Centre, University of Bradford.

TAN SW (1978) A Financial Model of North Sea Oil. MBA Thesis, Management Centre, University of Bradford.

TATLOW PK (1978) Local Authorities, Censorship and Public Communication: a Study of Decision-Making. MBA Thesis, Management Centre, University of Bradford.

THORNTON PC (1974) Consumerism in Britain. MSc Thesis, Management Centre, University of Bradford.

TIMSON D (1978) Resource Allocation by Linear Programming in a Further Education College. MBA Thesis, Management Centre, University of Bradford.

WALTERS IA (1977) Management in Education: Towards a Systematic Approach. MBA Thesis, Management Centre, University of Bradford.

WHITELAW M (1971) The Methodology of the Evaluation of Management Training. MSc Thesis, Management Centre, University of Bradford.

WILSON K (1974) Prospects and Problems of British Building Societies. MSc Thesis, Management Centre, University of Bradford.

WOODBRIDGE P (1974) The Importance of Businessmen's Motivations and Values in the Analysis of Corporate Objectives. MSc Thesis, Management Centre, University of Bradford.

YOUNG PM (1977) Risk Management in Local Government. MBA Thesis, Management Centre, University of Bradford.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci. Vol. 7, No. 4.
Pergamon Press Ltd 1979. Printed in Great Britain

ERRATUM

Omega, Volume 6, Number 3, 1978

The order of authors' names on p. 281 should read:

James W Gentry
TV O'Brien
(*November* 1977)

College of Business Administration
Oklahoma State University
Stillwater
Oklahoma 74074
USA

ol. 7
979

BLANK
PAGE

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 5, pp. 379 to 384
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1001-0379$02.00/0

EDITORIAL

Public Sector Decision-Making

INTRODUCTION

EVERYONE is aware that there is *something* rotten in the state of the United Kingdom as far as public expenditure and taxation are concerned. Arguably, public expenditure levels are too high: 'crowding out' resources which could otherwise be transferred to the private sector to rebuild investment capacity for a growth economy. Alternatively, public expenditure may itself be seen as an instrument of growth, especially where infant industries require initial capital to start up in a world of rapidly developing technology. The same school of thought might also point to the need for increased public expenditure to satisfy those wants and needs that have hitherto had to be sacrificed because of comparatively low wealth. This special issue of OMEGA is not, however, about the sheer size of the public sector, although the opening essay by Michael Common bears directly on it.

To set the scene, Common asks whether the traditional economic rationale for intervention by the state in a market economy holds fast in the light of repeated attacks by so-called 'liberal' economists. The substance of the attack is that the interventionists have exaggerated the degree of imperfection in the market—they have, as it were, found social costs under every stone. They have also, so the charge goes, overlooked the fact that intervention can itself be a social cost, showing up in infringements on the civil liberties and traditional freedoms of the individual.

Common has little time for either argument, and he is surely correct to point to the alarming intellectual paucity of some of the recent debates on external costs. This seems to have been conducted on the basis of trivial examples such as bee-keeping and apple growing, perhaps because those examples were the ones used in the seminal articles on social cost and which seem to have had such influence on intellectuals. The problem appears to lie in the absence of real-world awareness by the liberal critics of the severity of many externalities. Common points to the laws of thermodynamics and their impact on economic literature in terms of the pervasiveness they imply for waste mater-

ials discharge to the environment. Confining himself to this general area alone, he sees cause for considerable alarm. Moreover, it is an alarm that is not alleviated by arguments to the effect that sufferers and polluters can somehow 'get together' to bargain their way to some 'optimal' level of pollution. On the basis of some simple analytics he shows that the existence of external costs—costs to third parties which are not 'internalised' by the polluter—implies a positive price for the use of the environment to the firm and a zero price to consumers. Yet no market can operate with a price which is simultaneously zero and positive, so that market transactions are simply not a feasible option.

The 'liberal' case against infringement of civil rights is of course a powerful one. Quite how it links to notions of justice and fairness in public sector decisions is unclear, however. Moreover, free market principles are just as capable of infringing civil rights, as the growth of nuclear power amply demonstrates. There is no avoidance of the fact that more nuclear power, perhaps dictated by the free expression of public want, also means more surveillance of parties not directly party to the decisions to introduce that technology. There remains, then, a case for intervention and control. What the 'optimal' extent of intervention is, will remain a much debated issue. To be sure, it is more than the recent 'liberal economics' school of thought would accept on the basis of their misapplied logic, but perhaps less than others would call for.

Vol. 7
1979

RATIONALE FOR MAKING DECISIONS

The remainder of the essays in this volume deal with the issue of *how* to make decisions. Interestingly, many contributors draw attention to the apparent demise of cost–benefit analysis as the panacea for solving this complex issue. Certainly, cost–benefit (CBA) has experienced a severe decline in popularity. Some authors dwell on the reasons. Button in his essay points to the failure to 'monetise' the intangible items in a cost–benefit balance sheet. Lowe and Lewis draw attention to the considerable information problems of setting about a cost–benefit study, at least in the world of environmental problems. Nash criticises the procedures for establishing even the 'hard' estimates of benefits such as time-savings in transport projects, while Treddenick points out that even resource costs are not easy to measure if the cardinal rule that they reflect their opportunity cost is remembered. The question of 'who votes' is also raised. Many public investments now secure gains and costs for the current and immediate generations to come. Some impose costs on future generations, perhaps for many hundreds of years to come. How then are future generations' wants to be incorporated in a decision rule?

One view would be that since they will be wealthier than us they will have the time and resources to solve any inherited social cost problems. Such arguments from faith, or rather complacency, are difficult to accept. Simply because we have been used to what is effectively exponential growth over two centuries, in no way means that the future will be like the past. If future generations are likely to be worse off, there is an extra duty on current generations to leave them a capital stock while the 'stock' of inherited social costs should be minimised, at least in the sense of being made reversible if they lack the technology to deal with such problems. Moreover, since no-one knows what they will want, it makes more sense to plan on the basis of minimising their regret: not their regret at any level since that would impose an impossible decision-making framework on current generations, but perhaps minimising their maximum regret if they inherit problems on a scale larger than this generation forecasts. None of the essays in this volume tackles this complex issue, at least not directly. Lowe and Lewis do point to some of the strange effects that result from the use of positive discount rates such that long-term social cost problems can quickly 'disappear' with the use of present value criteria. While noting that positive discount rates can 'justify' the overuse of the environment's capacity to assimilate waste, they do not quite follow the issue through to note that physical laws will still operate in such a context such that 'catastrophic' results from overpollution are not thereby ruled out.

THE STATE OF COST–BENEFIT

Most contributors to this volume are satisfied that CBA must be revised to deal with the problems noted above, and we have stressed the intergenerational issue to add to their list of criticisms. This leads the authors to search for alternative measures. Barrett gives an excellent survey of the use of environmental impact statements as substitutes for CBAs, although, of course, the general way in which EISs are construed in the USA permits them to subsume CBA studies as well. There is much to be said for separating out the various impacts and showing them as clearly as possible so that they are not 'lost' in a single net present value figure. Equally, the dangers need emphasising since it is quite possible that the 'anti-CBA' pendulum has swung too far. To 'capture' all impacts in an EIS could simply be to engage in a wholesale and detailed description of a project. This may be no bad thing for single projects but the issue arises of how one *compares* two projects using EIS (or more than two). For the entries in an EIS matrix can be thought of as the components of any vector. Unless we have certain vector inequality, which is unlikely, we have no clear way of choosing between the two projects. Of course, this criticism may reflect an unwarranted desire to return to the days

where at least some of the 'judgement' in choosing is removed from the decision-maker. EIS advocates would point to the fact that two or more impact matrices would 'force' the decision-maker to enter into a judgement that is at least based on comprehensive information.

This is a valid point, although it holds for CBA as well. For much of the criticism of CBA has arisen because of the over-ambition of some of its proponents in claiming that monetisation was simply a matter of the 'state of the art': more and more research would bring about the correct answers, eventually. If, however, a humbler view is taken of CBA then it has its uses. For the disturbing feature about actual decision making, and several authors note it, is that it frequently takes place on *no rational grounds at all.* The standard response would be that all decisions are 'political' and cannot be taken on the basis of the kinds of criteria advocated by some of the authors in this volume—whether weighted indices of environmental impact (Lowe and Lewis), planning balance sheet approaches (Button), multi-criteria/multi-objective techniques, or whatever. But the 'political' argument is really little more than a mechanism for imposing constraints on the decision-maker such that the actual process of deciding is that much easier. Not that political decisions are unimportant. Decisions in the National Health Service may often have all too much to do with the special pleading of senior clinical personnel and all too little with an assessment of the opportunity cost of the requests they make.

Vol. 7
1979

Quite simply, then, there remains a strong case for employing any technique which forces a list of costs and benefits to be presented. Maybe impact statements and Delphi procedures offer this minimum condition. One could argue that CBA does likewise.

THE STATE OF THE ART IN PUBLIC DECISION MAKING

As the authors in this volume make clear, no single technique would appear to be applicable to the separate parts of the public sector. Gray and Steele see considerable benefits in the application of programme budgeting techniques but strike a note of some despair over the hostility to rational planning at all in the NHS. Treddenick notes the unavoidable nature of the use of cost-effective techniques in defence, but also sees problems in its application. Nash feels sufficiently disillusioned with CBA in the transportation field to want to argue for even less attempt at monetary evaluation than would be dictated by even apparent consensus on the 'hard' figures. Lowe and Lewis see little scope for monetary evaluation techniques in environmental decision-making.

One is left with the impression that the 'state of the art' is now one of a backlash against traditional CBA techniques and the existence

of a vacuum which analysts are busy trying to fill with the varied techniques outlined by Button in his useful survey article. It seems reasonable to argue, however, that none of these techniques is likely to have less difficulty than CBA, and, perhaps more tendentiously, one could suggest that fashionable critiques of CBA will be forced to give way as the pendulum swings back towards less ambitious but still useful cost–benefit frameworks. But that is a prediction and the test lies in the future.

PUBLIC PARTICIPATION

One other interesting development is noted in several of the essays. This is that the method of appraising projects cannot now be divorced from the issue of public participation in making those decisions. To this end one wonders again if an EIS fares any better than a CBA. Multi-objective techniques would certainly seem to fail on this count given their low index of comprehensibility, even for many experts. But it is right that public participation be stressed. The sheer speed at which technology is advancing has left many with a distinct feeling that they are alienated from the decision-making process. We can see the emergence of extra-parliamentary institutions as indirect evidence of exactly this. Public inquiries, Royal Commissions and so on, reflect the sheer incapacity of parliaments to deal with complex technological issues on a level that permits their detailed and careful assessment. For example, we are witnessing changes in the procedures for assessing road investments precisely because the objector to a motorway cannot understand the computerised techniques used for assessing traffic forecasts and time-saving benefits. Such a process can go too far: reduced to the lowest common denominator of understanding, one may get decisions based on no efficiency criteria at all, other than that participation has been maximised.

But if participation is to be one of the integral parts of any definition of efficient decision-making—and this seems as justified for persons affected by road schemes as it does for the patient whose decisions are made by proxy by his GP—then it all has to be linked to parliamentary scrutiny, at least for those decisions which, on one criterion or another, are judged of 'national' importance. Yet those links are tenuous in the current state of UK politics. No real mechanism exists for linking planning procedure at the local level with parliament, only with the relevant Secretary of State. Select Committees surely have a greater role to play and Heald, in his essay, surveys the political debate on this issue with admirable clarity. His prime concern, however, is with the overall capacity of parliaments to scrutinise public expenditure and taxation plans. The message is clear: those powers that do exist are distinctly inadequate and quite out of keeping with the workings of a sensible democracy.

CONCLUSIONS

In the space of one collection of papers it has not been possible to range across the many other issues that require discussion. It is hoped, nonetheless, that the papers in this special issue will serve to stimulate progress in thought on the three major areas identified: (i) the issue of just what the size of the public sector should be; (ii) the issue of how rationally to assess expenditure decisions within any given sized public sector; and (iii) the issue of how best to integrate the goal of public participation into the concept of efficient decision-making.

DAVID PEARCE
Guest Editor

Department of Political Economy
Edward Wright Building
University of Aberdeen
Dunbar Street
Old Aberdeen AB9 2TY
Scotland, UK

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 5, pp. 385 to 397
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1001-0385$02.00/0

External Costs: Myth or Rationale for State Intervention?

MS COMMON

University of Stirling, Scotland, UK

(*Received May* 1979)

The paper considers the extent to which problems of environmental pollution can be left to the market system. The concept of external cost is used to establish the presumption that market forces will lead to excessive pollution. The performance criterion, that of Pareto efficiency, leading to this presumption is discussed. It is shown that for significant problems of pollution it is not possible to limit the role of the state to that of re-defining property rights. What is necessary is state intervention by way of the taxing, or regulation, of discharges into the environment: where external effects are undepletable, markets cannot meet the requirements of Pareto efficiency. Government intervention to meet such requirements requires that the government has information which is, in practice, either unavailable or very costly. Intervention towards more restricted goals is discussed. It is shown that arguments to the effect that the role of the state can be restricted to defining private property rights are based on assumptions which are irrelevant to significant problems. In particular, it is argued that a recent publication by the Institute of Economic Affairs does not, as claimed, show either that external costs do not exist, or that they exist but imply a minor role for the state in the protection of the natural environment.

INTRODUCTION

A RECENT Institute of Economic Affairs publication [7] claims in its title that social, otherwise known as external, costs are a myth. Actually, a less strong claim is advanced in the publication: that problems of external cost exist, but can be solved by re-drawing the boundaries of private property rights so that market trading over external effects can occur. These are important claims. It is the existence of external costs which is the basis for an argument for government intervention in market systems. This argument is deployed primarily with respect to problems of environmental pollution: "the problem of environmental pollution is a simple matter of correcting a minor resource misallocation by means of pollution charges" [4]. The purpose of this paper is to assess the claims made by the Institute of Economic Affairs. For this purpose it is necessary to consider the nature of problems of environmental pollution, the concept of external cost and some recent developments in the economic analysis of environmental problems. It will become clear that while government intervention is necessary, it is not likely to 'solve' the problems completely: the resource misallocation to be corrected is not aptly described as 'minor', and the matter is hardly 'simple'. Too much can be claimed for governments, just as too much can be claimed for markets.

The nature of significant problems of environmental pollution

By way of definition, it will suffice to say that environmental pollution exists when the residuals arising in the activities of production and consumption are discharged into the natural environment in such quantities as to have adverse effects perceived by people. There is, again, no precise criterion which separates significant from non-significant examples of pollution, but precision is hardly necessary. The noise from a neighbour's lawnmower may be highly irritating, but the significance of this problem is an order of magnitude different

from that of what is generally meant by 'environmental pollution'.

It is now widely accepted that many pollution problems have global dimensions, though there exists disagreement on the precise nature of the problems. Thus, while some take the view that the buildup of CO_2 in the atmosphere will warm the planet and melt the ice-caps, others argue that particulates in the atmosphere are reducing the penetration of solar radiation with the prospect of leading to a new ice-age. An excellent exposition of these views, and other global aspects of pollution, is [21]: see also, for a more recent assessment of the situation, [13]. There also exist global problems associated with discharges of DDT, heavy metals and radioactive substances [21]: DDT residues have been found in whales which feed in the Chukchi and Bering Seas. Whatever the precise current state of knowledge on any of these questions, the possibility that the stability of the global ecosystem is threatened cannot be dismissed. Global pollution problems have three features of significance here. First, that the nature of the damage is uncertain but possibly catastrophic. Second, the impossibility of establishing a definite link between an observed problem (Bering Sea Whales) and specific sources of residuals emission (farms in Iowa or Japan). Third, that the origins and consequences transcend the boundaries of nation states. The problems are not, then, readily amenable to conventional economic analysis.

There remain cases of pollution which can reasonably be considered 'regional' in the sense that "the impact is largely constrained by some fairly definable environmental system that is less than global" [17]. Table 1 shows the quantities of gaseous residuals discharged into the atmosphere over the USA in 1971. The residuals listed all have adverse effects perceived by people, the most widely investigated being human health effects.

Taking, for example, the excess mortality induced by a 'typical' coal burning 1000 MW electricity plant, operating at a 75% load factory in the USA, the range of estimates is from 10 to 100 per year: see [14] and also [9]. For an analysis of the excess mortality implications of air pollution in the USA see [19]. It is a matter of every day experience, as well as an established research finding, that air pollution is especially a problem in large urban areas and cities. Definitionally, large cities contain large numbers of people affected by air pollution. It is probable, though not definitional, that large cities contain many sources of emission of residuals. Even if, in the absence say of private cars, there are but few sources, it is difficult to relate the emissions of a particular source to ambient pollution levels, and extremely difficult to relate those emissions to a particular individual's experience of ill-health or death. For large cities, the limiting case must be just one source of emissions into the atmosphere—the electricity plant say—with very many affected, in varying degrees perhaps, individuals. These remarks, it will be seen, are of central relevance to the possibility of leaving the control of air pollution to markets and private property rights. Obviously, similar remarks apply to other regional pollution problems such as the overuse of watersheds for residuals dumping.

It is not impossible to think of cases where there are few affected individuals, few sources, and clear links between source and damage: the noise of lawnmowers or hi-fi equipment in a residential area; a neighbour piling unsightly

Vol. 7
1979

TABLE I. ESTIMATED DISCHARGES OF GASEOUS RESIDUALS, NATIONWIDE BY SOURCE, 1971 (MILLION TONS)

Source	Carbon monoxide	Particulates	Sulphur dioxide	Hydrocarbons	Nitrogen oxides	Total
Transportation	77.5	1.0	1.0	14.7	11.2	105.4
Fuel combustion in stationary sources	1.0	6.5	26.3	0.3	10.2	44.3
Industrial processes	11.4	13.6	5.1	5.6	0.2	35.9
Solid residuals disposal	3.8	0.7	0.1	1.0	0.2	5.8
Miscellaneous	6.5	5.2	0.1	5.0	0.2	17.0
Total	100.2	27.0	32.6	26.6	22.0	208.4

Source: Council on Environmental Quality, *Environmental Quality—1973*, Washington, DC, 1973, p. 266.

rubbish in his garden; a workshop or farm that leaves rusting machinery about the place; a factory chimney in a remote and sparsely populated area. But such cases are hardly examples of the environmental pollution that is nowadays, rightly or wrongly, held to be a serious problem of recent origin. The significant pollution problems are characterised by many sources and/or many affected individuals, and by non-obvious links between source and individual damage.

THE SIMPLE ECONOMICS OF EXTERNAL COST

Consider some productive activity, associated with which are a number of desirable effects and a number of undesirable effects, where in both cases the magnitude of an effect increases with the level of the activity. Denoting the activity level by x, desirable effects as y's, and undesirable effects as z's:

$$y_i = y_i(x) \quad \frac{\partial y_i}{\partial x} > 0 \quad \text{for } i = 1,2,\ldots,m$$

$$z_j = z_j(x) \quad \frac{\partial z_j}{\partial x} > 0 \quad \text{for } j = 1,2,\ldots,n$$

The problem is to decide upon the level at which the activity should be operated. An obvious rule would be to run the activity at the level which yields the greatest excess of desirable over undesirable effects. This rule is nonoperational, as both the desirable and undesirable effects are heterogeneous so that neither class can be summed for comparison one with the other—apples cannot be added to pears, nor man-hours of labour to tons of fertilizer. The problem remains if there is just one effect of each class—apples are not commensurable with man-hours. An operational rule is to run the activity at the level which maximises the excess of benefit over cost. Benefit is the result of converting the measure of each desirable effect to that of value, by multiplying quantity by price, and summing the value of all desirable effects. Cost is the result of the same conversion and summation over all undesirable effects. Benefit and cost are value totals which can be compared one with another:

$$\text{Benefit} = \sum_{i=1}^{m} p_i y_i = B(x) \quad \frac{\partial B}{\partial x} > 0, \quad \frac{\partial^2 B}{\partial x^2} < 0.$$

$$\text{Cost} = \sum_{j=1}^{n} p_j z_j = C(x) \quad \frac{\partial C}{\partial x} > 0, \quad \frac{\partial^2 C}{\partial x^2} > 0.$$

The decision rule operate where the excess of benefit over cost is greatest can be differently expressed. From

$$\text{Maximise } \{B(x) - C(x)\}$$

we immediately get as necessary

$$\frac{\partial B}{\partial x} - \frac{\partial C}{\partial x} = 0$$

or

$$\frac{\partial B}{\partial x} = \frac{\partial C}{\partial x} \tag{1}$$

as giving the required level of operation for the activity. In this form the rule says: operate where marginal benefit equals marginal cost. Figure 1 shows graphically the equivalence of the two versions of the rule. In the marginal form, this rule, with terminological adjustments, is all pervasive in economics.

In either form, this rule solves the complex problem of deciding on the optimal level for an activity with many heterogeneous effects. It will be apparent that the role of prices in aggregation is crucial: different price sets will produce different levels for the activity and the associated effects, desirable and undesirable.

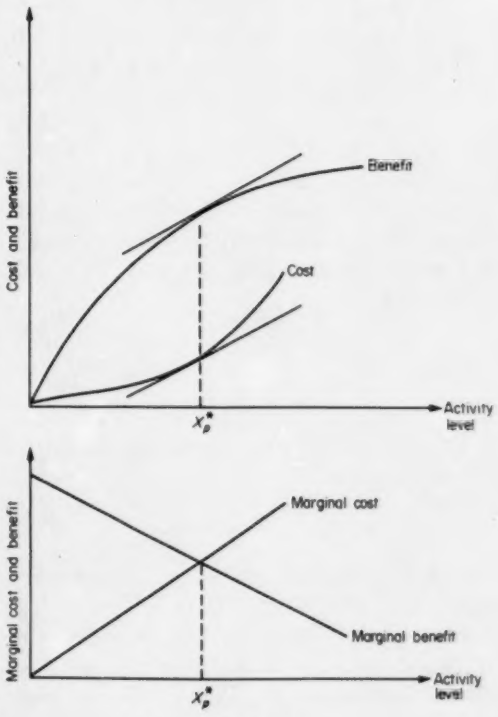


FIG. 1.

Consider the same productive activity to be looked at by two different decision units. One is the firm which actually operates the productive activity: its price set will be the set of market prices it faces. The other is some collective agency representing the interests of the society within which the firm exists. Under what circumstances will firm and agency compute the same optimal level for the activity? The answer is if firm and agency operate with identical price sets. Now, in some very closely specified circumstances, market prices will be the prices appropriate for use by the agency, perhaps more readily called the government, or the state. The nature and significance of these circumstances will be discussed shortly. For the moment, consider a case where there is one undesirable effect, the emission of a noxious residual into the atmosphere, to which the firm attaches a zero price where the government does not. Let this be the n'th effect, and take all other prices to be the same as between firm and government. Using the adjective 'private' to define the situation as perceived by the firm:

$$\text{Private Benefit} = \sum_{i=1}^{m} p_i y_i = PB(x)$$

$$\frac{\partial(PB)}{\partial x} > 0, \quad \frac{\partial^2(PB)}{\partial x^2} < 0$$

$$\text{Private Cost} = \sum_{j=1}^{n-1} p_j z_j = PC(x)$$

$$\frac{\partial(PC)}{\partial x} > 0, \quad \frac{\partial^2(PC)}{\partial x^2} > 0.$$

The optimal level of the activity for the firm is the level which maximises the difference between private benefit and private cost, this difference to be called here net private benefit, and often referred to as profit. Thus,

$$\text{Maximise } \{PB(x) - PC(x)\}$$

is the same as

$$\text{Maximise } \{NPB(x)\}$$

with the equivalent marginal form:

$$\frac{\partial(NPB)}{\partial x} = 0. \tag{2}$$

The firm's decision rule is: operate where marginal net private benefit is zero. Now, using the terms benefit and cost without qualification for matters looked at from the government perspective introduces a definition of external cost:

$$\begin{aligned}\text{Benefit} &= \text{Private Benefit}\\ \text{Cost} &= \text{Private Cost} + p_n z_n\\ &= \text{Private Cost} + \text{External Cost.}\end{aligned}$$

For the government, the optimal activity level is that which maximises:

$$\begin{aligned}\text{Benefit} - \text{Cost} &= \text{Private Benefit} - \text{Private Cost} - \text{External Cost}\\ &= \text{Net Private Benefit} - \text{External Cost.}\end{aligned}$$

The socially optimal level of the activity x is given by

$$\text{Maximise } \{NPB(x) - EC(x)\}$$

yielding the marginal version:

$$\frac{\partial(NPB)}{\partial x} = \frac{\partial(EC)}{\partial x}. \tag{3}$$

The social interest requires a level of operation such that marginal net private benefit equals marginal external cost. As shown in Fig. 2, the socially optimal level, X_s^*, given by (3) is lower than the optimal level for the firm, X_p^*, given by (2).

The question which now arises is: in what circumstances will the possibility of different price sets for firm and government not arise? Put slightly differently: under what circumstances will the pursuit of private gain through market transactions serve the interests of society as a whole? Economists have devoted a lot of attention to this question in the 200 years since Adam Smith [27] promoted the idea that market trading would reconcile the

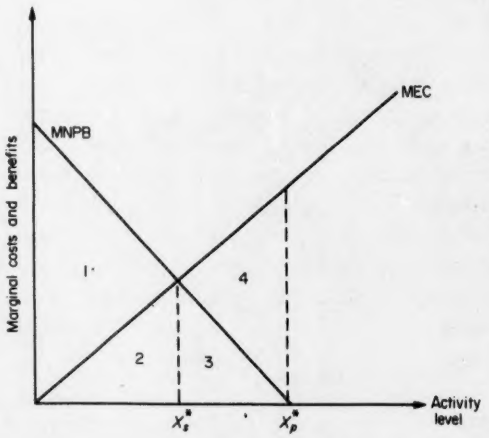


FIG. 2.

Vol. 7
1979

private and the social interest: the circumstances under which economic activity can be left to Smith's 'invisible hand' are now closely defined. The subject matter here is 'welfare economics' and 'general equilibrium analysis': a concise and rigorous account of the essentials, with references to the literature, is in Chapters 9 and 10 of [15]; see also [5]. For the present purposes, the most significant of the circumstances is one which is frequently left implicit: that there exist tradeable private property rights in all inputs to, and outputs from, productive activity. It is also required that all of the individuals in the economy be fully informed as to the consequences of any prospective market exchange. Also of relevance here is the condition that no participant in market exchange has the power to influence the terms on which exchange takes place: all traders are 'price-takers'. It will be clear that the conditions under which the invisible hand of the market promotes the social interest are restrictive. It is also the case that, given that the conditions are satisfied, the sense in which market trading maximises the social interest is rather restricted. The basic theorem of welfare economics says that, given the conditions described and some others of primarily technical interest, a market economy will secure an allocation of inputs to and outputs from production which is 'efficient' in the sense that no individual in the economy can be made better off except by making some other individual(s) worse off. This is not a strong result. It is that, under ideal conditions, all the potential gains from trade will be realised by self-seeking individuals. The condition of efficiency—often referred to as Pareto efficiency, or Pareto optimality—does not define a unique allocation. Consider, for example, two individuals shipwrecked on a desert island together with a limited amount of canned food. *Any* allocation of this food, including one in which one individual's share is zero, is Pareto efficient. Efficiency does not necessarily imply fairness. Given the condition that no individual can be made better off except by making some other individual worse off, there remains the ethical question of the distribution of 'well-offness' across individuals. The market is neutral in this respect. Given some pre-trade distribution of wealth, market trading will produce an efficient outcome. But, this is all that can be claimed for markets, even under ideal conditions. Given this, very few economists argue against government activity to achieve some view of fairness or justice in the distribution of wealth and income.

Indeed, the point of the basic theorems of welfare economics is that they indicate that the problems of distribution and efficiency can be approached separately: under the specified conditions, society can take and enforce a view on distribution, leaving the achievement of efficiency to the market system. This is convenient, for it means, if accepted as operative in actual economies, that economists can debate the need for government intervention to secure efficiency, without necessarily considering the distributive implications of intervention. These can be taken care of by re-distributive taxes and grants. Thus, the problem of external cost is a problem of the failure of the market to achieve efficiency. It is conventional to discuss proposed remedies only in terms of the achievement of efficiency, ignoring the distributional implications. This convention is followed, but not defended, here: for a discussion of some of the distributional issues which arise in the context of environmental protection see Chapter 13 of [2], and [11]. The separation of questions of efficiency and distribution does not mean, as is often implied, that the concept of Pareto efficiency is ethically neutral, or value free. As used, the concept requires that whether an individual is made better or worse off is decided only in terms of the subjective evaluation of the individual concerned. The value premise is then that the social good is to be measured only in terms of subjective individual preferences: for further discussion of this point see, for example, [25]. While it may be argued that in some circumstances this is innocuous, it will be clear from the previous section that it is not obviously appropriate when considering problems of environmental pollution.

In introducing the definition of external cost above, an example was used where a firm released, as an effect of its productive activity, a noxious residual into the atmosphere. To this effect, the firm attached a zero price where the interests of society as a whole, represented by the government, did not. The source of this example of market failure to secure efficiency in allocation is the absence of property rights

in the atmosphere. The atmosphere as recipient for residual discharge can be looked at as an input to production, or a particular atmospheric quality can be looked at as an output from production. Looked at either way, the atmosphere is not something in which there generally exist tradeable private property rights. As a result, the firm treats the atmosphere as a free input to production, or treats a polluted atmosphere as a consequence of its activities for which it is not made responsible. The result is overuse of the atmosphere: with respect to the efficiency criterion, the firm operates the activity x at too high a level and emits too much of the residual. If this is accepted, how can efficiency in allocation be achieved? In terms of Fig. 2, the question is how can the firm be induced to operate at the level X_s^* rather than at the level X_p^* where its own interest is best served? There are two sorts of approach to consider. The first involves changing the legal framework such that private property rights are created where their absence gave rise to the external cost problem. This solution involves government intervention of a limited kind: the government changes the rules of the market game, then withdraws. The second approach involves continual government intervention. The government effectively takes collective property rights, and it, or some appointed agency, then takes measure to induce the firm to operate at X_s^*. The most obvious form of 'inducement' is for the government simply to instruct the firm to operate at the level X_s^*. Such a regulatory approach is generally thought by economists to be less desirable than the imposition of taxation on residuals emission, such that the after-tax marginal net private benefit function cuts the horizontal axis at X_s^* in Fig. 2. Two points apply equally to the regulatory and taxation solutions. First, as already noted, they give the government a permanent role. Second, they require that the government, or its appointed agency, has the information necessary to identify the activity level X_s^*. This is, as critics of the collective property rights approach point out, a strong requirement, and one which the private property rights solution avoids since market bargaining will itself reveal where X_s^* is. Unfortunately, it is not possible to believe that the private property solution is applicable for significant pollution problems. In the case of air pollution used as an example here, the difficulties of establishing private property rights in the air over a large city, will, no doubt, have already struck the reader. As, no doubt, will the difficulties of conceiving what kinds of bargain would be struck, if private property rights could be established.

The next section of this paper will consider the case for, and nature of, the private property rights approach. The subsequent section will deal with the collective property rights approach. At this point it is necessary to point out that external costs are not the only possible source of market failure, and that external costs may occur together with other violations of the conditions necessary for markets to achieve efficiency, such as the existence of monopoly power. In such cases the analysis becomes much more complex, and leads to less clear cut implications than does the existence of external cost alone. Thus, for example, if the firm conducting the activity which gives rise to residuals emissions is a monopolist, it is not certain that the activity level it considers optimal, in the absence of any property rights in the atmosphere, will be too high for efficiency in allocation: it may be too low. These additional complexities are ignored throughout this paper: the interested reader is referred to [2], especially Chapters 6, 7 and 8.

Recognition of the existence of external costs goes back some way in the literature of economics. They appear, as do most ideas in economics, in Marshall's 'Principles' [22]. The first systematic treatment is by Pigou [26] in the 1920's: Pigou's proposed solution was a system of corrective taxes, and such are nowadays called 'Pigovian taxes'. With some exceptions (see [16]), the economics profession did not, until the early 70's regard external costs as a burning issue. The topic was not without theoretical interest, but the problem was not considered pressing in fact: for a useful review of the post World War II literature see [24]. This situation arose in part from an implicit endorsement of Pigou's view that external costs were exceptional, rather rare blemishes on the market system. This assessment has, at least with respect to environmental pollution, been drastically revised for many economists in the last decade. One factor in the re-assessment is a response to observed pollution problems. Another factor has been a revision in the view

Vol. 7
1979

of what it is that an economic system (any economic system) does. This revision is primarily associated with the materials balance principle, spelt out in [18]: a less formal, if more memorable, statement of the same perspective is Boulding's notion of the 'spaceship economy' [6]. According to the materials balance principle, an economy returns to the natural environment as residuals a quantity of materials approximately equal to the tonnage of raw materials extracted from the natural environment. Residuals discharges are then, routine and pervasive, not exceptional. In itself, this does not, of course, imply that external costs are pervasive. However, given the observed current lack of private property rights in airsheds and watercourses, there is a strong presumption that external cost problems are far from being rare or exceptional.

THE COASE THEOREM

An important, but sometimes misunderstood, contribution to the theoretical literature on external cost is Coase's 'The Problem of Social Cost' [8]. Published in 1960, this contribution has by now had conferred on it the status of theorem, which has two parts:

(i) in a world where no individual or firm can influence the terms on which trade takes place, where all individuals and firms have perfect information, and where private property rights can be costlessly established and enforced, external costs will not exist and private bargaining will secure an efficient allocation of resources.

(ii) in such a world, resource allocation will be efficient whatever the particular legal rules on the direction of liability.

Coase himself did not put matters quite this way, and it is possible to disagree about what he actually meant to say: see, for example, [24] and [20]. However, (i) and (ii) constitute what is now known as the Coase Theorem. It might be noted that (i) is effectively a restatement of the basic theorem of welfare economics: where they can, rational individuals will realise all potential gains from trade. It can also be noted that, with respect to (ii) especially, nothing is said about distribution.

The content of (i) and (ii) can be elucidated using Fig. 2, and extending the earlier discussion of the firm discharging a residual into the atmosphere. Take it that all of the emitted noxious residual falls eventually onto the lawns of an estate adjacent to the firm, damaging the grass, and that it is known to both firm and estate owner that the damage to the grass is due solely to the firm's emissions. Suppose that, by virtue of the enactment of a law giving individuals property rights in undamaged lawns, the estate owner is able to extract, costlessly, full compensation for damage from the firm. Then, estate owner and firm will reach a bargain which will involve the firm running the activity of production at the level X_s^* in Fig. 2. In that Figure, areas under the marginal external cost, MEC, and marginal net private benefit, MNPB, lines are value totals for damage costs borne by the estate owner and profits to the firm respectively. With a bargain involving the activity level X_s^*, the firm makes profits given by the area 1 plus 2, and pays out in compensation an amount given by the area 2. Consideration of marginal displacements to either side of X_s^* will indicate that this is the best that the firm can do for itself, given the liability rule of the legislation. The estate owner suffers damage given in value by the area 2, and receives full compensation. Bargaining itself reveals the efficient position. Now, suppose that a different law had been enacted, giving firms property rights in the atmosphere. In this situation it is open to the estate owner to seek reductions in the damage to his lawns by offering the firm compensation for curtailment of its activity. Again, assuming costless bargaining and equal bargaining power, the result of the estate owner so doing will be that the firm ends up operating at X_s^*. In this assignment of liability or property rights, X_s^* will involve the firm foregoing profits in amount given by the area 3, for which loss it receives full compensation from the estate owner, so that its final profit position is as it was before the legislation. The estate owner is better off than he was before the legislation: by paying out an amount given by area 3, he avoids damage which he values at the area 3 plus the area 4. Again, consideration of marginal displacements to either side of X_s^* will indicate that this is the best position for the estate owner, given this assignment of liability. The firm, in this case, is fully compensated for any lost profits. Again, bargaining itself reveals the efficient position.

Now, the assumption of costless bargaining is perhaps a reasonable approximation with one pollution generator and one pollution receiver. It may be a reasonable approximation where there are few generators and few receivers. It is clearly not a reasonable approximation where there are many generators and/or many receivers: in such cases bargaining, or transactions, costs must be non-zero and may be prohibitive. It follows that the Coase Theorem is simply irrelevant to large numbers cases. This is not solely a matter of transactions costs, as is sometimes implied. As already indicated here, significant problems in environmental pollution are almost certain to involve many affected individuals, no one of whom is able to identify the precise source of any damage felt. In such cases the problem is identifying an agent with whom to bargain: this problem too is sometimes swept under the transactions costs rubric. Bargaining requires, on the part of the pollution affected individual, the perception of damage at a point in time such that it is possible to secure a reduction in damage: with, for example, air pollution over large cities, such perception is simply not possible. Large numbers on either the generating on the receiving side necessarily mean high bargaining, or transactions costs, and are typically associated with highly imperfect information. The position on the Coase Theorem is well put in [2]:

> "It thus seems to us that the role of voluntary negotiation among individual decision makers is of limited applicability for environmental policy. Once asserted, the point is almost obvious. Yet it is important to nail it down. For example, it is surprising how many casual readers of the literature one encounters who feel that Coase's shafts have dealt a fatal blow to the Pigovian solution in practice because, if negotiation moves resource allocation toward the ideal, the imposition of the optimal Pigovian taxes on top of this will prove too much of a good thing. The argument, so far as it goes, is valid. But if, in most important externalities problems, negotiation is impractical and virtually nonexistent, the damage to the Pigovian position inflicted by this point can hardly be very serious."

UNDEPLETABLE EXTERNALITIES AND PIGOVIAN TAXES

It is necessary now to outline briefly an analysis which is more appropriate to significant problems of pollution. This section of the paper draws heavily on the important recent (1975) work of Baumol and Oates [2]. Given a concern with large numbers cases, it is appropriate to switch from the partial equilibrium analysis of previous sections, to a general equilibrium approach, looking directly at utility functions and production functions rather than at benefits and costs. An individual's utility function is a description of his preference system, relating alternative consumption patterns to an (ordinal) index of the individual's welfare, or utility. A firm's production function is a description of its technology, relating input combinations to outputs. An externality, or external effect, is present whenever some individual's utility, or firm's production, function includes real (i.e. non-monetary) variables, the levels taken by which are chosen by others without particular attention to the effect on the welfare of the individual or firm in question. On this definition, externalities can be beneficial as well as detrimental. Also, firms can give rise to externalities affecting individuals and/or firms, and individuals can give rise to externalities affecting firms and/or individuals. Here there will be considered only detrimental externalities generated by firms, that is, in the previous terminology, cases of external costs generated by firms. As pointed out by Bator [26], many externalities have the dimension of a public good, or bad, in that they are non-rival in consumption. Thus, all the residents of city experience the same air quality when they use the streets: nor does an increase in one individual's respiration rate decrease the pollution available to the other city residents. Clearly, such non-rivalness in consumption is the appropriate first approximation to most, if not all, significant examples of environmental pollution. To avoid some of the connotations of the term 'public goods', Baumol and Oates have called externalities with this characteristic 'undepletable' externalities. Depletable externalities can be defined, but it is difficult to think of interesting examples of detrimental externalities which are depletable in the pollution context. For discussion of this point, and of the taxonomy of externalities generally, see Chapters 3 and 4 of [2].

Consider now an economy with two individuals, A and B, each consuming two commodities, X_1 and X_2, each of which is produced by a firm using labour supplied by the individuals. In producing output, each firm generates

Vol. 7
1979

and emits into the atmosphere 'smoke': the level of smoke in the atmosphere adversely affects each individual and each firm's output for given labour input and smoke emissions. The externality is undepletable. It should be noted that the use of a case with two firms and two individuals here is solely to simplify the exposition notationally: the analysis extends to any number of individuals, commodities and firms (see Chapter 4 of [2]), and is not concerned with the small numbers case discussed in the previous section. In an obvious notation, the problem of defining a Pareto efficient allocation can naturally be set up as:

Maximise $U^A(X_1^A, X_2^A, L^A, Z)$;

$$u_1^A > 0, \quad u_2^A > 0, \quad u_L^A < 0, \quad u_z^A < 0$$

subject to $U^B(X_1^B, X_2^B, L^B, Z) = \bar{U}^B$:

$$u_1^B > 0, \quad u_2^B > 0, \quad u_L^B < 0, \quad u_Z^B < 0$$

$X_1^A + X_1^B = X_1(L_1, S_1, Z)$:

$$x_{1L} > 0, \quad x_{1S} > 0, \quad x_{1z} < 0$$

$X_2^A + X_2^B = X_2(L_2, S_2 Z)$:

$$x_{2L} > 0, \quad x_{2S} > 0, \quad x_{2Z} < 0$$

$$L_1 + L_2 = L^A + L^B$$

where $Z = S_1 + S_2$ gives the ambient pollution level, and where

$$u_1^A = \partial u^A/\partial X_1^A, \quad u_z^A = \partial u^A/\partial Z \text{ etc}$$

$$x_{1L} = \partial X_1/\partial L_1, \quad x_{2S} = \partial X_2/\partial S_2 \text{ etc.}$$

Form the Lagrangian function with multipliers: λ on the constraint that B_s' utility be held constant; μ_1 on the constraint for the production of commodity I; μ_2 on the constraint for the production of commodity 2; μ_3 on the labour availability constraint. Then, the problem specified has first order conditions which can be written as:

$$\left.\begin{aligned} \frac{u_1^A}{u_2^A} = \frac{u_1^B}{u_2^B} &= \frac{\mu_1}{\mu_2} \\ \frac{u_1^A}{u_L^A} = \frac{u_1^B}{u_L^B} &= \frac{-\mu_1}{\mu_2} \end{aligned}\right\} \tag{4}$$

$$\mu_1 x_{1L} - \mu_3 = 0, \quad \mu_2 x_{2L} - \mu_3 = 0 \tag{5}$$

$$\left.\begin{aligned} u_Z^A + \lambda u_Z^B + \mu_1(x_{1S} + x_{1Z}) + \mu_2 x_{2Z} &= 0 \\ u_Z^A + \lambda u_Z^B + \mu_1 x_{1Z} + \mu_2(x_{2Z} + x_{2S}) &= 0 \end{aligned}\right\}. \tag{6}$$

Consider now utility maximisation by households and profit maximisation by firms, where commodity prices are μ_1 and μ_2, where the wage rate is μ_3, and where there is tax of t per unit on emissions of smoke. For household A, the problem is

Maximise $U^A(X_1^A, X_2^A, L^A, Z)$

$$\text{subject to } \mu_3 L^A - \mu_1 X_1^A - \mu_2 X_2^A = 0$$

which gives first order conditions:

$$\frac{u_1^A}{u_2^A} = \frac{\mu_1}{\mu_2}, \quad \frac{u_1^A}{u_L^A} = \frac{-\mu_1}{\mu_3}. \tag{7}$$

Considering household B in the same way will give:

$$\frac{u_1^B}{u_2^B} = \frac{\mu_1}{\mu_2}, \quad \frac{u_1^B}{u_L^B} = \frac{-\mu_1}{\mu_3}. \tag{7'}$$

Utility maximisation by households satisfies the necessary conditions (4) for Pareto efficiency, as (7) and (7') correspond to (4). Looking at profit maximisation by the firm producing commodity 1, the problem is

$$\text{Maximise } \mu_1 X_1(L_1, S_1, Z) - \mu_3 L_1 - t S_1$$

which has first order conditions:

$$\mu_1 x_{1L} - \mu_3 = 0, \quad \mu_1 x_{1S} - t = 0.$$

Considering the profit maximising production of commodity 2 in the same way, the conditions describing profit maximisation in production are:

$$\mu_1 x_{1L} - \mu_3 = 0, \quad \mu_2 x_{2L} - \mu_3 = 0 \tag{8}$$

$$\left.\begin{aligned} -t + \mu_1 x_{1S} &= 0 \\ -t + \mu_2 x_{2S} &= 0 \end{aligned}\right\} \tag{9}$$

Comparing (5) and (8), they are seen to be the same. Comparing (6) and (9), they are seen to be the same if and only if

$$t = -\{(u_Z^A + \lambda u_Z^B) + \mu_1 x_{1Z} + \mu_2 x_{2Z}\} \tag{10}$$

Thus, it is shown that given the imposition of a uniform tax on smoke emissions, at the rate given by (10), the market will meet the conditions for efficiency in allocation.

It is very important to note that what has been shown is that efficiency requires *only* a tax on emissions. There is *not* required, as is sometimes argued, any payment to the individuals and firms who are affected by the externality. Looked at slightly differently, the point is that efficiency requires an asymmetry in that externality generators are made to face a price, but the externality recipients are not: from the point of view of the generating firm smoke carries a price, from the point of view of the suffer-

ing individuals and firms smoke does not carry a price. Now, no price determined in any kind of market can exhibit such asymmetry: no 'price' can be simultaneously zero and non-zero. With undepletable externalities there is no possibility of securing efficiency by establishing some new market. What is required is a Pigovian tax, and the government is the only agency capable of levying this. However, what (10) further shows is the nature of the information requirements which face a government seeking to secure efficiency in allocation in an undepletable externality situation. From (10), it is apparent that the government needs to know the utility functions of individuals and the production functions of firms. Further, what the government needs to know is not the marginal utilities and marginal products evaluated at the consumption and output levels obtaining prior to its intervention, but the marginal utilities and products evaluated at the efficient consumption and output levels. This implies a knowledge not just of preferences and technologies in the neighbourhood of the pre-intervention situation, but of preferences and technologies over their whole domain. This is an unrealistic requirement to place upon governments.

For this, and other reasons (such as the possibility that the existence of external effects may itself lead to the violation of the conditions under which the existence of an efficient competitive equilibrium can be proved—see Chapter 8 of [2]) some economists have concluded that the goal of efficiency in allocation, in the face of externalities, has to be abandoned. Is there, in this case, anything useful for economists to say? The so-called 'least-cost' theorem states that uniform taxes per unit of a particular residual discharge are the cheapest way of securing any arbitrary reduction in the level of pollution: for a full account of this theorem see [3] or Chapter 10 of [2]. Here 'arbitrary' means only that the reduction required has not been established with reference to the criterion of efficiency. Suppose that on the best medical and scientific advice it is necessary to cut ambient pollution by 10%. Then, the cheapest way to proceed is to levy a uniform tax per unit of discharge across all discharge sources, rather than to require a 10% reduction by all sources. The essential point is that the cost of achieving a 10% reduction in discharge must be expected to vary across sources, i.e. firms. A uniform tax rate means that firms will each cut back on discharges by an amount which reflects their assessment of the post-tax costs and benefits. The overall reduction will be concentrated with those firms where reduction is relatively cheap. Further, this is achieved without the need for the government, or its agency, to know anything about the firms' production functions, and the tax rate required to achieve the overall 10% reduction can be discovered by an iterative trial and error process. The conditions required for the least cost theorem are fairly weak. However, the theorem's wide applicability is achieved at the cost of abandoning the goal of allocative efficiency. It should be noted especially that the least cost theorem does not say that the costs of achieving the desired pollution standard will necessarily be small, it says only that, as measured by market prices they will be least. What is essentially a variant of the "uniform tax to achieve an arbitrary standard" proposal, but with a different assignment of property rights, is Dales' proposal [10] for the auction of pollution rights. For further discussion of the least cost theorem, and a description of empirical studies of particular environmental quality management problems see [17], which provides further references to the literature.

THE MYTH OF SOCIAL COST?

The preceding sections of this paper have shown that with respect to significant environmental pollution, there is a presumption that there is a necessary role for on-going state intervention. As noted at the beginning of this paper, this position is attacked in a recent publication of the Institute of Economic Affairs [7], entitled *The Myth of Social Cost* by SNS Cheung, with contributions by CK Rowley and J Burton. It is now possible to assess that attack. If some of the remarks which follow are rather strongly expressed, it must be borne in mind that the authors concerned have explicitly addressed themselves to a non-specialist audience on a politically sensitive topic, claiming particular technical expertise.

In his Prologue, Rowley begins equivocally: "Society *might* be better off if the 'problem' of

Vol. 7
1979

social cost had never been *discovered*" (italics added). Rowley is not prepared to deny the existence of social, i.e. external, costs. Nor is he sure that society would benefit from ignorance of their existence. While claiming that Coase [8] refuted the case for government intervention where external costs exist, Rowley notes that Coase's theorem assumes zero transaction costs, that "Transaction costs may frequently increase with the number of individuals", and that "the transactions costs of resolving an externality problem may be high". Indeed, "the existence of a problem of social cost in itself indicates the presence of 'transactions' costs" for "Why else would individuals forgo the gains from exchange which always exist where there are externalities?" Thus, Rowley has to concede that there are circumstances where "government intervention will appear appropriate". The concession is followed by a series of warnings: governments are not perfect, governments are not costless, governments respond to voters and pressure groups. With these warnings in mind, equivocation can be abandoned: "The *supposed* existence of 'social cost' has been one of the foremost *pretexts* for which such (individual) freedom has been transgressed and by which the authority of government has been extended" (italics added). No evidence is offered for this remarkable empirical assertion. If by 'social cost' Rowley means here, as he does elsewhere, an external cost associated with environmental pollution, the assertion is not remotely plausible. Such extravagance promotes neither the rational analysis of government behaviour nor the cause of individual freedom, which is confronted by more real and serious threats than government intervention in markets for the purpose of reducing environmental pollution.

Burton's contribution takes the form of an Epilogue, which adds little of substance and requires only two comments here. The first concerns Burton's exposition of the Coase Theorem, where all references are to a singular externality 'producer' and a singular externality 'consumer'. Such precise reporting does not prevent Burton from concluding that: "The Coase analysis knocked the supports from under the Pigovian analysis, and pointed to entirely different implications for public policy". As will now be clear, such 'entirely different implications' are relevant to a restricted set of problems which do not include the significant environmental pollution problems. Secondly, some comment is necessary on Burton's discussion of the 'problem of the commons' with respect to the recent problems—mass starvation—of the Sahelian region of Africa. Burton's point of departure is the assertion that private, as opposed to communal, land ownership is necessary and sufficient to prevent desertification. This assertion is unfounded. It is now a well known result in the theory of renewable resources that sole ownership is not sufficient to prevent the extinction of a renewable resource [28]. The American dustbowl was, for example, created under institutions of private, not communal, land ownership. Further, private ownership is not a necessary condition: societies where private ownership of natural resources is unknown have successfully exploited such over millenia. Burton himself notes: "For millenia... the peoples of this (Sahelian) belt have utilised a system of tribal or communal ownership of land". Notwithstanding this interesting fact, Burton asserts that the basic cause of desertification is "the attenuation of private property rights", the problem so arising being "exacerbated by government attempts to increase pastoral productivity". Burton confidently dismisses as either irrelevant or harmful all measures other than the institution of a 'private property solution'. Despite the nonexistence of private property rights for 'millenia', Burton finds it necessary to offer no remarks on how to move towards this 'indispensable' solution. It would, presumably, involve some government intervention. Rowley's concern with the threat to individual freedom from taxes on discharges of harmful residuals stands in interesting contrast to Burton's blithe indifference concerning the implications, in Africa, of imposing upon a culture what is, by his own account, an alien system of land tenure.

The longest contribution to *The Myth of Social Cost* is that of Cheung. This contribution is less polemical than the other two, but it is not, as is claimed by the Institute's Editorial Director, "up-to-date in the light of the latest developments in the debate between economists". Cheung takes as the definitive statement of the case for state intervention the work of Pigou [26] published in 1920: the most recent reference cited on the intervention

side of the debate is [23] from 1952! Most of Cheung's contribution is taken up by a simple numerical example, used to illustrate various ways in which the partial equilibrium analysis of the external cost problem can be presented. Recent general equilibrium treatments, such as those in the previous section of this paper, are ignored completely. The contribution of Coase is discussed at some length. Cheung is actually very clear on the status of this contribution: "Under the preceding broad definition of transaction costs, and accepting the conclusion of the Coase theorem, to say that transaction costs are zero is identical to declaring the existence of *private* property rights" (italics in the original). The Coase theorem works by assuming away the problem.

The interesting part of Cheung's contribution is his section on "Fallacies in Reporting Observations", which is presumably seen as the empirical basis for the assertion, in the title of the paper, that social costs are a myth. Cheung claims that "Using imaginary 'facts' to support imaginary policies seems habitual in the Pigovian tradition". Cheung considers two cases where divergences between private and social price sets have been suggested. Pigou [26] took the view that tenant farming was less efficient than owner farming: Cheung provides what he takes to be a refutation based on data for China for 1920–1935. Meade [23] illustrated the possibility of the existence of an external benefit the relationship between beekeeping and apple farming: Cheung provides evidence for the USA to show that "contractual arrangements have long been routine between farmers and beekeepers". This is the totality of the empirical evidence offered by Cheung on the existence of external effects. Not one significant problem of environmental pollution is mentioned here. Indeed, Cheung states: "The costs of private contracting are not always and necessarily lower than the costs of deriving and enforcing public policies to achieve the same end. What is true of the bees should not be generalised to all other economic activities".

Cheung warns that "any government action can be speciously justified on efficiency grounds by the simple expedient of assuming that transaction costs in the market are high and that the costs of governmental control are low". Readers of *The Myth of Social Cost* should also be warned that it is possible to construct specious arguments against government action by extrapolating results which hold only for special cases, the Coase Theorem being a case in point. While the three authors of the Institute of Economic Affairs are generally, but not always, careful to avoid analytical errors and to qualify their technical statements, they do lend their names to what is a misleading and mischievous overall exercise. According to the Institute's Editorial Director: "If our three authors are right, the continued teaching in Britain of the simplistic conventional approach to social cost/benefit is seriously flawed and dangerously misleading". This assessment is in no way warranted by the three authors' actual contributions. Rather, it is the Institute of Economic Affairs which misleads by presenting as a balanced and up to date account of the literature, what is an unconvincing attack on an outdated strawman using an irrelevant weapon. There do exist significant problems of environmental pollution which require state action: neither the pretence that they do not exist, nor repeated incantations of the name of Coase, will make them go away. It is, of course, equally foolish to pretend that government action, though necessary, can cause the problems to vanish. There are no panaceas: "The truth, however unpleasant, is that the most we can do is to prevent any unnecessary depletion of resources and any unnecessary deterioration of the environment, but without claiming that we know the precise meaning of unnecessary in this context" [12].

Vol. 7
1979

REFERENCES

1. Bator FM (1958) The anatomy of market failure. *Q. Jl Econ.* **LXXII**, 351–379.
2. Baumol WJ & Oates WE (1975) *The Theory of Environmental Policy*. Prentice-Hall, Englewood Cliffs, New Jersey, USA.
3. Baumol WJ & Oates WE (1971) The use of standards and prices for protection of the environment. *Swed. Jl Econ.* **LXXIII**, 42–54.
4. Beckerman W (1972) Economists, scientists and environmental catastrophe. *Oxf. Econ. Pap.*, 327–344.
5. Bohm P (1973) *Social Efficiency: a Concise Introduction to Welfare Economics*. Macmillan, London, UK.
6. Boulding K (1966) The economics of the coming spaceship earth. In *Environmental Quality in a Growing Economy* (Ed. Jarrett H) John Hopkins Press, Baltimore, USA.
7. Cheung SNS, Rowley CK & Burton J (1978) *The myth of social cost*. Hobart Paper **82**. Institute of Economic Affairs, London, UK.

8. COASE R (1960) The problem of social cost. *Jl Law & Econ.* **III,** 1–44.
9. COMMON MS & PEIRSON J (1979) Environmental benefits and the economics of UK solar energy schemes. *Energy Econ.* not yet publ.
10. DALES JH (1968) *Pollution, Property and Prices.* University of Toronto Press, Toronto, Canada.
11. DORFMAN R (1977) Incidence of benefits and costs of environmental programs. *Am. Econ. Rev. Pap. & Proc.* **67**'(1), 333–340.
12. GEORGESCU-ROEGEN N (1975) Energy and economic myths. *Sth. Econ. Jl* **41,** 347–381.
13. GRIBBEN J (1979) Woodman spare that tree. *New Scient.* **81,** 1016–1018.
14. HAMILTON LD & MORRIS SC (1974) *Health effects of fossil fuel power plants.* Proc. Eighth Midyr Top. Symp. of the Health Physics Soc. on Popln. Exposure. Knoxville, Tennessee, USA.
15. INTRILIGATOR MD (1971) *Mathematical Optimization and Economic Theory.* Prentice–Hall, Englewood Cliffs, New Jersey, USA.
16. KAPP KW (1950) *The Social Costs of Private Enterprise.* Spokesman, Nottingham. Revised, extended and republished as *The Social Costs of Business Enterprise.* Spokesman, Nottingham, UK, 1963 and 1978.
17. KNEESE AV (1977) *Economics and the Environment.* Penguin Books, London, UK.
18. KNEESE AV, AYRES RU & D'ARGE RC 1970) *Economics and the Environment: a Materials Balance Approach.* John Hopkins Press, Baltimore, USA.
19. LAVE LB & SESKIN EP (1973) An analysis of the association between US mortality and air pollution. *Jl Am. Statist. Assoc.* **68,** 284–290.
20. LIN S (ed.) (1976) *Theory and Measurement of Economic Externalities.* Academic Press, New York, USA. (see especially chapters by Zerbe and Auten).
21. *Man's impact on the global environment: assessment and recommendations for action.* (1970) Rep. of the Study of Critical Envir. Probl., MIT Press, Cambridge, Massachusetts, USA.
22. MARHSALL A (1930) *Principles of Economics.* 8th ed., Macmillan, London, UK.
23. MEADE JE (1952) External economies and diseconomies in a competitive situation. *Econ. Jl,* 54–67.
24. MISHAN EJ (1971) The postwar literature on externalities, an interpretive essay. *Jl Econ. Lit.* **IX,** 1–28.
25. NASH C, PEARCE D & STANLEY J (1975) An evaluation of cost–benefit analysis criteria. *Scott. Jl Polit. Econ.* **XXII,** 121–134.
26. PIGOU AC (1920) *The Economics of Welfare.* Macmillan, London, UK.
27. SMITH A (1776) *The Wealth of Nations.*
28. SMITH VL (1977) Control theory applied to natural and environmental resources: an exposition. *Jl Envir. Econ. & Mgmt,* **4,** 1–24.

ADDRESS FOR CORRESPONDENCE: *MS Common, Department of Economics, University of Stirling, Stirling, FK9 4LA, Scotland, UK.*

BLANK
PAGE

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 5, pp. 399 to 409
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1001-0399$02.00/0

Models for Decision-Making in the Public Sector

KJ BUTTON

Loughborough University, UK

(*Received May* 1979)

The public sector is composed of a heterogeneous selection of undertakings which, for a variety of economic, political and historic reasons, are under direct government control. The considerable differences between these undertakings, combined with a paucity of common characteristics, has resulted in a variety of decision-making techniques being employed. In many cases there is an affinity with private sector firms, and techniques technically akin to the profit maximisation model are adopted although social welfare considerations supplant notions of financial returns. As with the private sector, textbook models are frequently abandoned to make way for more pragmatic criteria when informational or computational problems become severe. Additional to these undertakings using techniques parallel to those of the private sector, there are others for which specific decision-making models have been developed. These are mainly in areas where effective demand is considered inappropriate as a method of allocation and appraisal. Finally, there is some evidence that, as the multi-dimensional nature of public sector decisions becomes more widely accepted, the notion of satisficing is gradually gaining acceptance replacing that of maximising.

INTRODUCTION

THE POST-WAR period has witnessed a substantial growth in the public sector as government has deemed it desirable, for a variety of economic, social and political reasons, to play an increasing rôle in the provision of goods and services. Whereas at one time the meaning of economic efficiency was reasonably clear, the growth of an extensive public sector, where cost minimisation and profit maximisation are less compelling objectives, has led to the adoption of much wider concepts of efficiency. This does not mean that the financial considerations which play such an important rôle in private sector decision-making are entirely displaced, but rather that they are supplemented and modified by public-sector undertakings to reflect a much wider and less tangible set of goals and objectives.

The extent to which public sector decision-making deviates from models of the private sector varies quite considerably between different areas of government involvement. Quite clearly, in some sectors, such as education, health, police, defence and the provision of roads, the lack of a direct charging mechanism renders commercial criteria useless. At the other extreme, many of the nationalised industries operate in a market situation and, although controls are exercised to prevent monopoly exploitation, financial criteria have a very important function in decision making.[1] As we see later these differing features of public sector undertakings provide a useful method of categorisation when discussing decision-making frameworks.

The differing characteristics of public sector activities should not be taken as a reflection

[1] The 1978 White Paper on the nationalised industries, for example, argues that, "An adequate level of nationalised industry profits is essential to the continual well being of the industries and their customers and the economy as a whole. They provide some of the funds for the very large investment programmes necessary to maintain supplies and services to the public". (26, pp. 22–3).

of differing objectives. Rather they indicate differences in the constraints confronting the decision maker. The overriding objective which underlies all public sector decision-making is the maximisation of society's welfare but the varying characteristics and differing operational environments of the sector's undertakings present a complex of constraints which cover social, political and environmental considerations as well as those of a financial nature. Education and health, for example are deemed to be 'merit goods' [21] for which no charge should be made, while the provision of transport infrastructure involves taking account of a variety of external costs, such as environmental damage, which do not normally enter into market trading [41]. Even for commodities which are bought and sold, the government takes cognizance of a much wider range of constraints than would be the case with private sector provision. Decisions concerning energy, for example, and the exploitation of coal, gas and oil reserves involve taking a very long term view of the nation's needs in addition to potential short term profits.

Although there is a single objective in the overall policy formulation, public sector undertakings require, for pragmatic reasons, to specify more concrete goals by which to gauge their performance.[2] The operational goals of different sectors vary quite substantially as emphasis is placed upon different components of the overall social welfare function. It is the differences in constraints and goals which has, until comparatively recently, prevented systematic methods of decision-making being introduced into public sector undertakings. Until the mid 1960s, decision-making in the public sector was characterised by *ad hoc* rules of thumb. Concern with the allocation of public expenditures in general, as expressed in the Plowden Report of 1961 [23], led to attempts at improving decision-making throughout the public sector.[3]

It was realised that the traditional emphasis on inputs (i.e. the money, man-hours etc. consumed by projects) did not sufficiently relate expenditures to the goals which were being pursued. Further, most decisions tended to be based upon short term considerations and, specifically, on a single year's costings which seemed inappropriate for many public sector undertakings with very long term policy horizons. Such criticisms led to the gradual acceptance, in broad terms, of the planning-programming-budgeting systems (PPBS) framework for decision-making [2, 3]. The United Kingdom was not unique in this and mirrored a trend already begun in the United States and also being pursued at a slightly slower and more measured pace in Canada, Sweden and France [53].

Vol. 7
1979

PLANNING-PROGRAMMING-BUDGETING SYSTEMS

One writer on PPBS once said "there seems to be an unspoken 'gentleman's agreement' that the basic terms need never be defined" [19] and certainly discussions of the approach are beset with terminological ambiguities. Even now no consensus on the appropriate jargon can be taken as definite. Rather than add to the confusion by introducing new terminology or attempting to unravel the existing tangle, this paper broadly accepts the definitions used by Foster [15]. His jargon certainly is not always consistent with other sources but it does at least provide an acceptable and uniform basis upon which to discuss PPBS.[4]

In essence PPBS is a framework within which attempts are made to identify expenditures and social outlays as closely as possible, and to relate them, in opportunity cost terms, to specific goals. Ideally, there would be one programme of action corresponding to each of the goals of the public sector. In practice, the real world involves taking into account administrative and physical considerations which prevent this ideal being attained. Government goals, for instance, usually span several of its departments which makes strategic decision making extremely difficult. Nevertheless, the overall Public Expenditure Survey forward

[2] The problem of defining an acceptable social welfare function has exercised economists for many years; see [14].

[3] Perhaps the most tangible recognition of this need was the establishment of public expenditure review procedures in 1969 [25]. For details of the development of public expenditure management see [17].

[4] PPBS is itself often called Programme Budgeting or Output Budgeting although at other times these simply refer to stages in the wider process. System analysis is another term which is often introduced and can take several meanings, sometimes with specific connotations related to operations research but frequently of a more general nature.

budget in the UK is based upon a functional classification in the PPBS manner. Despite this gradual movement, PPBS is still not employed with effect at the strategic level within the UK but it is widely used for tactical decision-making within departments. (Strategic decisions follow from inter-departmental analysis, co-ordinated by the Public Expenditure Survey Committee, and are made collectively by Ministers early each year.) The hope is that with the aid of PPBS procedures, a department may more clearly define what its objectives are, what activities contribute to these objectives, what resources or inputs are contributing to these objectives and what is actually being achieved or what the outputs are. PPBS can be seen, therefore, as a method of organising material, marshalling facts and ideas together, in such a way that better decisions are reached. It is not in itself a method of arriving at the 'correct' decision.

Broadly, PPBS involves the integration of three basic functions. Planning is concerned with surveying the existing situation, projecting future trends and determining appropriate goals and objectives. Programming involves formulating various projects and courses of action to achieve the agreed set of goals and objectives. Budgeting incorporates both costing (in broadest economic sense) the projects and actions to be implemented, and the establishment of a system of controls to monitor these costs. At the tactical level several forms of PPBS are employed but these can usefully be divided into two main groups which correspond roughly to the types of output of the sectors involved.

In general terms it is possible to categorise certain public-sector authorities as providing a tangible output which is, or could be, marketed: in other words, goods and services to which normal neo-classical economic demand analysis may be applied. The nationalised industries, transport, water and local authority housing are obvious examples. Such activities may be appraised using techniques akin to those employed by private sector undertakings. The only real difference is that while private undertakings generally attempt to maximise producers' surplus (the difference between selling price and cost) public sector undertaking in this categorisation tries to maximise combined producers' plus consumers' surplus. The latter reflects the welfare enjoyed by consumers in excess of the prices they pay and may be measured by the area under the demand curve above price [9].

In contrast, there are many public sector activities for which, for one reason or another, demand is difficult to determine and, consequently, benefits are virtually impossible to assess. Such sectors often provide merit goods, where 'need' [58] is thought a more relevant criteria than demand, and, also, frequently have complex distributional issues to consider. Health, defence, police services, and the arts, broadly fall into this category. Additional to these two main groupings, which form the focus for the remainder of the paper, there exists a *pot pourri* of other public expenditures which act as instruments of government policy. A whole range of activities come under this umbrella including housing grants, subsidies to nationalised industries, social security, regional location grants etc. but they can be distinguished from the other categories by virtue of the fact that the government is not directly supplying goods or services. Public sector expenditure of this kind is only indirectly related to resource utilisation [4]. The decision-making framework used for expenditures in this category differs from situation to situation but, because the incidence of costs and benefits tends to be important, political considerations often dominate any economic analysis.

THE QUASI-COMMERCIAL SECTOR

The standard method of decision making applied to those public sector undertakings where demand is thought the relevant criteria upon which to allocate resources, is cost–benefit analysis (CBA).[5] The underlying model is closely allied to that of perfect competition in the traditional economic abstractions of the private firm. Rather than profit maximisation as the objective, however, the aim is to select

[5] A vast literature describing and discussing standard theoretical CBA techniques has grown up over the years (possibly justifying a CBA of its own worth). Good accounts may be found in [10, 32, 42, 44, 54] while [1] provides a number of case studies.

the course of action which maximises net benefits, in the very broadest sense, adjusting all costs and benefits to correspond to those that would prevail in a perfectly competitive market. The pure form of CBA was summed up over a decade ago in a well known definition found in the classic article by Prest and Turvey [50, p. 683]; "CBA is a practical way of assessing the desirability of projects, where it is important to take a long view (in the sense of looking at repercussions in the further as well as the nearer future) and a wide view (in the sense of allowing for side effects of many kinds on many persons, industries, regions, etc.)—i.e. it implies the enumeration and evaluation of all the relevant costs and benefits". The unique procedure for policy appraisal implied by Prest and Turvey has gradually widened out, and CBA has now become rather more of a generic title for a collection of related techniques rather than the name of one specific set of calculations. In its original form, CBA superficially resembles the discounting procedures employed by large private companies and involves estimating the net present value of a project or plan by finding:

$$\sum_n \sum_i \left[\frac{P(a_i B_{ni}) - P(b_i B_{ni})}{(1+r)^n} \right]$$

where:

$P(a_i B_{ni})$ = the probable benefit expressed in monetary terms to be gained by individual (or groups of individuals) i in year n as the result of the project's completion. The benefit B_{ni}, is given a weight of a_i to reflect the importance attached to i's welfare;

$P(b_i C_{ni})$ = the probable cost in monetary terms to individual i in year n associated with the project. C_{ni} is given a weighting of b_i to reflect the importance attached to the costs borne by i.

$(1+r)^{-n}$ = the relative weight attached to a cost or benefit occurring in year n. The form of the weight reflects the decreasing importance attached to the more distant attributes of the project.

If the net present value is positive then the plan is accepted. When there are several options, but mutual exclusiveness limits acceptance to one, then selection is based upon the highest incremental net present value (with different sizes of alternative investments, the larger is accepted if the additional discounted benefits generated exceed the additional discounted costs). In the case of options of the same scale the comparison of benefit–cost ratios may be preferred on the gounds of simplicity.

In the late 1960s, this comprehensive, all embracing CBA framework was accepted by many as the panacea to all the problems of decision-making in the public sector. It was taken up with particular enthusiasm by the transport sector where it was seen to have potential, not simply as a method of large scale investment appraisal (as typified by the studies of the M1 Motorway [6], the Victoria Line Underground Railway [16], the Channel Tunnel [7], London's system of ringway urban motorways [18] and the siting of a Third London Airport [8]), but also as a standard method of routine decision-making (e.g. to assess railway social service subsidies [40] and inter-urban road investments [12]). The 1968 Town and County Planning Act, with its emphasis on 'strategic' rather than physical planning, also encourages the adoption of CBA for local level land-use planning decision-making. In 1967 the concept of a test discount rate (at that time, 8%) was introduced to encourage consistency in the temporal weighting of costs and benefits in different areas of public sector involvement and "to ensure that the calls of the public and private sectors upon resources do not get out of line with each other over the long term" [24, p. 7].

The initial enthusiasm for CBA has, however, waned in recent years and the all embracing 'textbook' model set out above has been superseded in many areas by a variety of pragmatic adaptations. The reasons for this change of attitude are multifarious. At the most basic level there is general mistrust of any procedure which appears to offer a simple, numerical answer to difficult and delicate decisions which, by their nature, involve a variety of subjective considerations. There is a suspicion that CBA can be manipulated to provide justifications for specific lines of action corresponding to the

Vol. 7
1979

view of the practitioner [51].[6] It has become apparent that there is no single, 'proper' way to perform a CBA but rather a number of alternative approaches, each corresponding to a particular set of moral notions [43]. These were problems explicitly recognised by early theoreticians in the field but their suggested qualifications and reservations were not always carried into practice. The need for decision-makers using a CBA framework to make their assumptions explicit is now widely accepted and the quasi-mechanical procedure implied by the equation set down above is now treated as a starting point of a much wider evaluation of different courses of action.

More specific problems with the pure CBA model stem from putting the underlying welfare economic theory into practice. There are often considerable difficulties in physically tracing out and measuring all of the consequences of a project and, even when measurement is feasible, it is frequently almost impossible to place anything more sophisticated than arbitrary monetary values on such effects.[7] In fields where research work on evaluation has been long standing (e.g. the valuation of time, life and noise), the techniques developed rely upon estimating the hypothetical market prices for intangles which would prevail if they were traded in competitive conditions. Evaluations depend, therefore, on the willingness of the affected parties to pay and government dictates and priorities are considered inappropriate.[8] The partial nature of CBA makes this approach particularly unsuitable for appraising large scale schemes or projects where wide ranging multiplier repercussions on the economy as a whole can affect the constellation of relative prices and production which form the basis against which intangibles are evaluated.

The weighting schemes required in CBA procedures are particularly difficult to devise. The distributional weights reflecting the impact of a decision on different sectors of the community pose specific problems of inter-personal comparisons which must, by their nature, be subjective.[9] In addition, although the test discount rate (and, more recently the recommended rate of return [26]) represents an attempt at introducing standard temporal weights across areas of public sector activity, the rates used are somewhat arbitrary and only partially reflect the social opportunity cost of resources. The actual estimation of the time stream of costs and benefits is itself difficult and the introduction of sensitivity and other, usually cruder, techniques to reflect the stochastic nature of the forecasts introduces further elements of subjectivity.

This lengthy list of criticisms levelled against the pure CBA approach, and the obvious defects of the technique as a practical tool for assisting in public sector decision-making in all but special cases, has led to a number of variants of the basic model being developed. The nature of these modifications depend, in the main, on the relative importance attached to the various constraints confronting the decision-makers in the individual sectors, and upon the types of goals envisaged as appropriate for the kind of activity being considered. In transport, for example, the problem of placing acceptable values on all intangible costs of road investments has resulted in the employment of a two stage procedure. COBA, a computerised package, evaluates those effects which are comparatively easy to quantify and forecast, and where evaluation methods are relatively well advanced (i.e. changes in vehicle operating

[6] The basic underlying problem with CBA is well summed up by Wildavsky [57, p. 297], "Although cost–benefit analysis presumably results in efficiency by adding the most to national income, it is shot through with political and social value choices and surrounded by uncertainties and difficulties of computation".

[7] An up-to-date collection of some papers setting down different approaches to the evaluation of intangible and externalities is found in [48].

[8] In practice, priorities and dictates do enter into the estimation procedure. An example is time evaluation; the transport ministry has, in the past, used an average value of time savings for all travellers whereas there is ample evidence that time valuation is positively correlated with the income group under consideration. The justification for the practice was to prevent undue bias being given to schemes likely to be utilized by high income groups.

[9] Early CBA studies excluded such weights, being simply interested in aggregate costs and benefits. They were founded on the so-called Kaldor compensation principle [30]. This essentially says that if the strict Pareto criteria cannot be met (i.e. it is impossible to benefit one person without making another worse off) then a weaker practical test is to see whether it would be *possible* for the losers to compensate the gainers. No actual compensation need take place, however; compensation just has to be potential. The work of Little [39] in the 1950s has slowly permeated CBA practice and with it questions of equality have supplemented those of efficiency. Little advocates the need to both meet a Kaldor type test and provide evidence that the project improves income distribution.

costs, journal travel times and accident rates). Such results, appropriately discounted, are then subjected to discussion at public enquiry. It is here that wider social, environmental and distributional questions can be raised. The difficulty of comparability between the quasi-scientific evidence offered by COBA and the unquantified, verbal account of the wider social issues of a road investment, however, causes serious disquiet [13]. A similar problem arises when it is impossible to assess the longer term effects of a course of action, again comprehensiveness must be sacrificed. In the case of the London urban motorway proposals, for example, the sheer complexity of tracing out the temporal pattern of interacting transport demands over a complicated road system resulted in the calculation of only a first year rate of return.

In other areas, the problems of evaluating intangible costs or forecasting the long term implications of a project are not the main causes for concern. Rather there are difficulties in assessing the monetary value of benefits; a situation not unusual in public sector undertakings whose actions have wide ranging impacts throughout the economy. Under such circumstances it is often common to view the decision-making process in terms of cost-effectiveness.[10] Klarman *et al.* [31, p. 49] provide a useful definition, "Cost-effectiveness, rather than cost–benefit, is employed when various benefits are difficult to measure or when the several benefits that are measured cannot be rendered commensurate. Under cost-effectiveness analysis costs are calculated and compared for alternative ways of achieving a specific set of results". In such cases it must be possible to set goals in physical terms so that the exercise becomes one of selecting that course of action offering the lowest opportunity cost method of meeting these goals. Hence projects are assessed in terms of cost per unit of output.

[10] There is some problem with terminology here but cost-effectiveness may be likened to cost-minimisation in the private sector where some predetermined output level has been decided upon independently. Trade-off analysis involves more abstract goals than cost-effectiveness (e.g. in the case of transport it may be time savings or reduced accidents while cost-effectiveness concentrates upon physical goals such as four lane motorway mileage) and is a more refined, though not necessarily practical, variant of it.

The Central Electricity Generating Board (CEGB) uses a cost-effectiveness approach when deciding upon investment policies with regard to new generating capacity. The future demand for electricity is determined and the plant accepted which offers the greatest discounted cost saving over the expense of making more use of the oldest existing capacity. Where a single goal of this kind can be specified the technique has practical qualities, but in most cases there are secondary goals (e.g. in the case of CEGB safety is clearly important) which cannot easily be ignored. In these cases there are serious trade-off problems in selecting priorities amongst the goals. Further, the ideal cost-effectiveness study employs measures of opportunity cost which often present difficulties as severe as those encountered when trying to quantify and evaluate benefits. For these reasons cost-effectiveness analysis is most helpful in deciding upon the most efficient way of implementing a prior made decision. In the case of the CEGB, for instance, it is stated policy to have sufficient capacity to meet demand. With the location of a Third London Airport (which was strictly a cost-effectiveness study) [8] it had already been decided a new airport was required and only the question of the siting was at issue.

THE SOCIAL SECTOR

The 'merit' nature of services provided by health, defence and education authorities, combined with the distributional and public goods (i.e. goods where the exclusion principle cannot apply) problems associated with these sectors, precludes the practical application of CBA techniques. In particular, the idea of evaluating intangibles in terms of ability to pay is felt to be especially inappropriate. In their place output budgeting is employed with goal orientated accounts as the primary basis for analysis. The technique was initially adopted by the Ministry of Defence for the Defence Policy Reviews in 1964 as 'functional costing' and soon taken up by the Department of Education and Science [11] and the Home Office, concerning police in particular [55], with the Department of Health and Social Security [29] following in 1971. The technique essentially concerns allocating resources within a sector in such a way that further changes in this allocation would

Vol. 7
1979

only reduce society's wellbeing. The allocation is, therefore, within the sector a Pareto optimal one. It has been pointed out [4] that this necessarily assumes both that the sector in question is fairly clearly delineated, with resource consumption well represented by public outlays in the sector, and that output categories may be clearly and discretely defined without problems of cost allocation. While the first condition is quite well satisfied by most of the government sectors employing the technique, output categorisation does pose problems with respect to the second criteria.

Some sectors suffer from the problem of relating outputs to goals; outputs often overlap several goals and vice versa. In defence, for instance, the programme budget is presented in terms of output expenditures, rather than commitments or goals, because of, firstly, the multipurpose nature of the forces retained and, secondly, the complex relationship between general support and front-line activities. Education is an area where it is easier to relate outlays to specific groups (i.e. nursery education, higher education, education for the 16–19 year olds, postgraduate education etc.), but again it is impossible to relate these outlays to the underlying goals of education (e.g. social cohesion, higher labour productivity etc.). The degree of interdependence between different outputs and different goals, the problem of separating means from ends, makes it impossible in practice to allocate costs in the required fashion.

While the programme budget approach permits an assessment of the relative costs of switching resources between different activities and outputs, specific issues arising within defence, education and the like are often subjected to more detailed 'special studies'. Special studies form an important part of the programme budgeting approach of these sectors and usually involve the application of either CBA or cost-effectiveness procedures. The complexity of attempting to utilise these techniques in the educational, defence and health fields increases the resource costs of decision making substantially and, in general, this type of treatment is either reserved for topics where difficulties of quantification are less severe or where the need to make the appropriate decision is central to the overall success of the department's policy. In the case of health, for example, cost-effectiveness techniques have been applied to assessing the desirability of alternative treatment mixes offering identical results [31, 49], similarly there have been CBA studies of educational projects [59]. Basically, these special studies are employed in order to select from a set of alternative actions each leading to a previously specified goal.

MULTI-CRITERIA APPROACHES

The multifacet nature of the social welfare function which public sector undertakings are concerned with maximising, together with the constraints confronting management in these sectors, underlie the problems of decision-making. With those sectors taking demand as the parameter for benefit measurement, CBA techniques provide only a partial solution to the problem, while the socially orientated sectors have difficulties in weighting their numerous goals. In the face of these difficulties the need for new approaches to decision-making is becoming widely accepted. The chairman of British Rail points, for example, to the need for an approach which "can be understood by ordinary intelligent people... incorporates the methods of analysis developed by welfare economists over the last decade or so,... gets away from the naive position adopted by the early cost–benefit man which seemed to imply that every consideration could be perfectly weighted and that therefore there was a single best solution" [47].[11] The Leitch Report on trunk road investment takes a similar stance, "the right approach is through a comprehensive framework which embraces all the factors and groups of people involved in scheme assessment" [13, p. 133]. There is also some discussion over the need for such changes at the local authority decision-making level [22].

Multi-criteria decision making techniques move away from the idea of maximising and are more akin to the management theories of satisficing [52]. Rather than attempt to seek

[11] The railways encounter particular difficulties in making investment decisions. Part of British Rail's operations (e.g. inter-city and freight services) are committed to straightforward commercial criteria while other services are deemed socially necessary and are assessed using CBA type techniques. Allocation of resources between these two types of activities poses serious problems because of the incompatability, except in highly stylised circumstances, of the different types of return calculated [20].

optimum solutions, which, for practical reasons, are likely to be unobtainable, the decision-maker selects actions complying with a range of criteria which describe minimally satisfactory alternatives. In the case of education, for example, one seeks options which provide certain set levels of literacy, numeracy etc. rather than alternatives which maximise some goals, say literacy, at the expense of others.

Decision-making models based upon this type of framework are still in their infancy, certainly as far as practical applications are concerned. Land-use planning and regional policy analysis are exceptions to this situation and a variety of multi-criteria techniques have been developed and tested in these areas [46]. It seems probable that some of these methods may have potential for wider application in other areas of public sector involvement.

The planning balance sheet (PBS), initially developed by Lichfield for town planning purposes[12], is similar to CBA in that all costs and benefits are included and that distributional considerations are not neglected, but it is a movement in the direction of multi-criteria analysis in that not all the effects of the various schemes are translated into monetary terms. Where evaluation is difficult, physical values are used, and when quantification is impossible, ordinal indices or scales may be employed. A socio-economic account is drawn up setting out the full effects of each course of action indicating the extent to which various groups in the community will be affected. These are then compared to the predetermined planning goals which are selected as reflective of community needs. Alternative plans are ranked under each objective heading using ordinal ranking procedures and the ranks are then added together to produce a ranking of the plans with respect to the objectives taken as a whole—the scheme with the lowest algebraic total being deemed the best. A modified version of this type of approach, the project impact matrix, has recently gained the favour of the Leitch Committee [13].

[12] The PBS approach is essentially pragmatic in nature and, rather than evolving from a theoretical base, is an applied offshoot of CBA, developed and polished over a series of applications [34, 35, 36, 37, 38].

[13] A criticism of the approach and, in particular, the implied cardinal ranking is contained in [33].

The PBS attempt to extend CBA into a multi-dimension framework, however, requires substantial data inputs and only offers a partial solution to the problem of making inter-personal comparisons.[13] The ranking-process has to reflect social preferences which are themselves difficult to ascertain and, even if a consensus is possible, the ordinal nature of the ranking suggests a loss of efficiency in the technique. Although value judgements cannot be avoided, the PBS approach has the merit that, unlike the pure CBA model, these are made explicit rather than hidden in a final single NPV figure. The construction of the initial socio-economic account can often, in itself, be educational and shed considerable light on salient questions the decision-maker should be asking.

The PBS method falls between CBA and multi-criteria decision-making methods, containing elements of both the maximisation principles underlying CBA but without the emphasis on monetary evaluation of all the projects' effects. Multi-criteria decision-making techniques generally involve introducing weights to reflect the relative priorities attached to the various outcomes associated with different causes of action. Although these weights may be ordinal in nature, they are more often cardinal being based upon available empirical evidence. A number of multi-criteria approaches have been devised each attempting to achieve a multi-dimensional compromise between the wide diversity of goals and costs which are embodied in public sector choice. The approaches differ in their method of presentation, the level of mathematical sophistication involved and the amount of data input required. Several of the techniques rely upon geometrical representation to produce multi-dimensional scalings [5] while others involve a considerable degree of intuition. In general, however, these particular techniques tend to be rather specialised in nature and are only of practical use in certain specific and unusual circumstances. Of more practical value are some of the weighting techniques which already enjoy a degree of practical acceptance and for which the theory is comparatively well advanced.

The introduction of weightings permit the effects of a projected action to be reduced to a single, summary figure offering the possibility

Vol. 7
1979

of direct comparisons with other options. Variations on this theme abound.[14] The goal-achievements matrix approach [27], for example, offers an explicit treatment of various goals and applies a set of predetermined weights to them so that each option can be assessed in terms of goal achievement. To facilitate this, the goals are related to physical measures to reflect the extent to which they have been achieved. The final goal achievement account employs the weighted index of goal achievement to determine the preferred course of action. Another variant, concordance analysis [45], offers pair-wise comparisons of alternatives by estimating two indices; the first, the concordance index, reflecting the weighted degree by which one project dominates the other in terms of the goals met and the second, the disconcordance index, showing the weighted degree to which it is itself dominated.

To date, much of the work in the field of multi-criteria decision-making has focussed on devising mathematically consistent methods of handling weights and basic data. The major practical problem, however, is in devising the weighting scheme to be adopted; CBA may be seen as a special case of the multi-criteria approach in this context, with monetary values being employed as weights. Despite the practical difficulties of evaluating intangibles and the problems of equality, the CBA method is possibly more readily accepted than some of the alternative schemes suggested for multi-criteria procedures. The idea of adopting the implicit weighting scheme of past decisions [56], for instance, has the desirable merit of assisting in the continuity of decision-making but, even if a practical method of excavating such weights from the past was devised, the implicit assumption that the motivations behind past decisions were correct and socially desirable is open to question. Further, past decisions were based upon social proprieties and morals of the period in which they were made and these are unlikely to have held constant. Indeed, since public sector decisions have very long term implications, past actions are even less likely to reflect the priorities of future generations. Suggestions that the weights could be derived from political debate, say in the House of Commons, begs the question of quantification and assumes full information on the part of speakers. A further proposal that hypothetical questioning of individuals about imaginary decisions will reveal preferences [28] has an academic appeal, but the problem of ensuring unbiased responses, and the doubtful general applicability of the answers obtained, suggests severe practical limitations.

CONCLUSIONS

The very nature of the goods and services provided by the public sector renders the profit maximising models widely employed to describe private sector decision-making inapplicable. This is not to say that in certain areas of public sector activity there is not an objective function similar to that generally associated with the private sector but rather that the perception of the public sector is much wider. The models underlying CBA and cost-effectiveness analysis have their counterparts in the standard economic theories of profit maximising and cost minimising but the costs and the benefits considered are much more defuse and the view taken, much longer. Recent developments at the theoretical level, together with the views of some management and advisory bodies, suggest that there are movements in the public sector to shift away from the traditional maximising/minimising approaches to decision-making and towards the acceptance of a set of much weaker, but more pragmatic, approaches. In particular, the interest shown by British Rail, local authorities and road investment agencies in multi-criteria assessment techniques indicates that notions of satisficing are replacing those of maximising. It seems very doubtful, however, whether with our current state of knowledge, multi-criteria assessment can, in the short term, do more than make some of the assumptions underlying existing methods more transparent. The emphasis on specifying the exact objectives being pursued and the setting out of the repercussions of decisions may, however, in the longer term induce a greater public involvement in public sector decision-making and, with this, the necessary information of society's preference function may be revealed. Certainly,

[14] See Nijkamp [46] for a wide ranging survey of the various multicriteria techniques which have been developed although few have to date been employed in actual decision-making exercises.

the complexity of many existing public sector decision-making aids tends to discourage social involvement.

REFERENCES

1. BARKER PJ & BUTTON KJ (1975) *Case Studies in Cost Benefit Analysis.* Heinemann, London, UK.
2. BRIDGEMAN JM (1969) Planning programming budgeting systems I. *O. & M. Bull.* **24,** 167–177.
3. BRIDGEMAN JM (1969) Planning programming budgeting systems II. *O. & M. Bull.* **25,** 16–27.
4. BYATT IC (1977) Theoretical issues in expenditure decisions. In *Public Expenditure: Allocation Between Competing Ends* (Ed. POSNER M). Cambridge University Press, Cambridge, UK.
5. CLARKE D & RIVETT BHP (1978) A structural mapping approach to complex decision making. *Jl Ops Res. Soc.* **29,** 113–138.
6. COBURN TM, BEESLEY ME & REYNOLDS DJ (1960) The London–Birmingham motorway, traffic and economics. *Road Res. Lab. Pap.* **46,** HMSO, Crowthorne, UK.
7. COOPER & LYBRAND ASSOCIATES (1973) *The Channel Tunnel: a United Kingdom transport cost benefit study.* HMSO, London, UK.
8. *Commission on the Third London Airport* (1971). Report. HMSO, London UK.
9. CURRIE JM, MURPHY JA & SCHMITZ A (1971) The concept of economic surplus and its use in economic analysis. *Econ. Jl,* **81,** 741–799.
10. DASGUPTA AK & PEARCE DW (1972) *Cost–Benefit Analysis: Theory and Practice.* Macmillan, London, UK.
11. DEPARTMENT OF EDUCATION & SCIENCE (1970) *Output budgeting for the Department of Education & Science.* Ed. Plann. Pap. **1,** HMSO. London, UK.
12. DEPARTMENT OF THE ENVIRONMENT (1972) *Getting the best roads for our money: the COBA method of appraisal.* HMSO, London, UK.
13. DEPARTMENT OF TRANSPORT (1978) *Report of the advisory committee on trunk road assessment.* HMSO, London, UK.
14. FOSTER CD (1966) Social welfare functions in cost benefit analysis. In *Operational Research and the Social Sciences* (Ed. LAWRENCE J). Tavistock, London, UK.
15. FOSTER .CD (1971) *Politics, Finance and the Role of Economics: an Essay on the Control of Public Expenditure.* Allen & Unwin, London, UK.
16. FOSTER CD & BEESLEY ME (1963) Estimating the social benefit of constructing an underground railway in London. *Jl Res. Statist. Soc.* Series A, **126,** 46–56.
17. GOLDMAN S (1973) *The developing system of public expenditure management and control.* Civil Service College Studies, **2,** HMSO, London, UK.
18. GREATER LONDON COUNCIL (1970) *GLDP inquiry proof: transport.* Proof E12/1, Greater London Council, London, UK.
19. GROSS BM (1969) The new systems budgeting. *Pub. Adm. Rev.* **29,** 116–132.
20. HARRISON AJ & MACKIE PJ (1973) *The comparability of cost benefit and financial rates of return.* Govt Econ. Serv. Occ. Pap. **5,** HMSO, London, UK.
21. HEAD JG (1966) On merit goods. *Finanzarchiv,* **25,** 1–29.
22. HEADICAR PG & COOPER JSL (1978) Relating actions to intentions. In *Proceedings of PTRC Annual Summer Meeting of the University of Warwick Seminar K.* PTRC, London, UK.
23. HM TREASURY (1961) *Control of public expenditure.* Cmnd **1432.** HMSO, London, UK.
24. HM TREASURY (1967) *Nationalised industries: a review of economic and financial objectives.* Cmnd **3437.** HMSO, London, UK.
25. HM TREASURY (1969) *Public expenditure—a new presentation.* Cmnd **4017.** HMSL.
26. HM TREASURY (1978) *The nationalised industries.* Cmnd **7131.** HMSO, London, UK.
27. HILL M (1968) A goal achievement matrix for evaluating alternative plans. *Jl Am. Inst. Planners,* **34,** 19–29.
28. HOINVILLE G (1971) Evaluating community preferences. *Envir. & Plann.* **3,** 33–50.
29. HURST JW (1977) Rationalising social expenditure—health and social services. In *Public Expenditure: Allocation Between Competing Ends* (Ed. POSNER M). Cambridge University Press, Cambridge, UK.
30. KALDOR N (1939) Welfare propositions and interpersonal comparisons of utility. *Econ. Jl,* **49,** 549–52.
31. KLARMAN HE, FRANCIS J'S & ROSENTHAL GD (1968) Cost effectiveness analysis applied to the treatment of chronic renal disease. *Med. Care,* **6,** 48–54.
32. LAYARD R (Ed.) (1972) *Cost—Benefit Analysis.* Penguin, Harmondsworth, UK.
33. LEAN W (1967) Economic studies and assessment of town development. *Jl Town Plan. Inst.* **53,** 148–152.
34. LICHFIELD N (1966) *Cost–benefit analysis in town planning: a case study of Cambridge.* Cambridgeshire & Isle of Ely County Council, UK.
35. LICHFIELD N (1966) Cost–benefit analysis in town planning: a case study of Swanley. *Urb. Stud.* **3,** 215–250.
36. LICHFIELD N (1969) Cost–benefit analysis in urban expansion—a case study: Peterborough. *Region. Stud.* **3,** 123–155.
37. LICHFIELD N & CHAPMAN H (1968) Cost benefit analysis and road proposals for a shopping centre—a case study: Edgeware. *Jl Transp. Ec. Pol.* **2,** 280–320.
38. LICHFIELD N & CHAPMAN H (1970) Cost benefit analysis in urban expansion—a case study: Ipswich. *Urb. Stud.* **7,** 153–189.
39. LITTLE IMD (1950) *A Critique of Welfare Economics.* Oxford University Press, Oxford, UK.
40. MINISTRY OF TRANSPORT (1969) *The Cambrian Coast Line.* HMSO, London, UK.
41. MISHAN EJ (1971) The postwar literature on externalities. *Jl Econ. Lit.* **9,** 1–28.
42. MISHAN EJ (1971) *Cost–Benefit Analysis, An Informal Introduction.* Allen & Unwin, London, UK.
43. NASH C, PEARCE DW & STANLEY J (1975) An evaluation of cost–benefit analysis criteria. *Scott. Jl Polit. Econ.* **22,** 121–134.
44. NEWTON T. (1972) *Cost–Benefit Analysis in Administration.* Allen & Unwin, London, UK.
45. NIJKAMP P (1975) A multicriteria analysis for project evaluation: economic–ecological evaluation of a land reclamation project. *Pap. Reg. Sci. Assoc.* **35,** 87–111.
46. NIJKAMP P & VAN DELFT A (1977) *Multi-Criteria Analysis and Regional Decision Making.* Martinus Nijhoff, Leiden, Netherlands.
47. PARKER P (1978) *A way to run a railway.* Haldane Memorial Lecture.
48. PEARCE DW (Ed.) (1978) *The Valuation of Social Cost.* Allen & Unwin, London, UK.
49. PIACHAUD D & WEDDELL JM (1972) The economics of treating varicose veins. *Int. Jl Epidem.* **1,** 287–294.
50. PREST AR & TURVEY R (1965) Cost–benefit analysis: a survey. *Econ. Jl,* **75,** 683–735.
51. SELF P (1975) *Econocrats and the Policy Process: the Politics and Philosophy of Cost Benefit Analysis.* Macmillan, London, UK.

Vol. 7
1979

52. SIMON HA (1959) Theories of decision-making in economics and behavioural science. *Am. Econ. Rev.* **49**, 253–83.
53. SUDAMA T (1977) PPBS and theories of decision making, bureaucracy and politics. *Publ. Fin.* **32**, 354–373.
54. SUGDEN R & WILLIAMS A (1978) *The Principles of Practical Cost-Benefit Analysis.* Oxford University Press, Oxford, UK.
55. WASSERMAN GJ (1970) Planning, programming, budgeting in the police service in England and Wales. *O. & M. Bull.* **25**, 197–210.
56. WEISBROD BA (1968) Income redistribution effects and benefit-cost analysis. In *Problems of Public Expenditure Analysis* (Ed. CHASE SB) Brookings Institute, Washington, USA.
57. WILDAVSKY A (1966) The political economy of efficiency: cost–benefit analysis, systems analysis and program budgeting. *Publ. Adm. Rev.* **26**, 292–310.
58. WILLIAMS A (1974) 'Need' as a demand concept (with special reference to health). In *Economic Policies and Social Goals* (Ed. CULVER AJ) Martin Robertson, London, UK.
59. WISEMAN J (1965) Cost-benefit analysis in education. *Sth. Econ. Jl*, **32**, 1–14.

ADDRESS FOR CORRESPONDENCE: *K Button, Esq, Department of Economics, Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK.*

BLANK
PAGE

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 5, pp. 411 to 419
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1001-0411$02.00/0

Cost-Benefit Analysis — a Critique

RK TURNER

University of East Anglia, Norwich

(*Received May* 1979)

This paper surveys some of the main criticisms of cost–benefit analysis (CBA) that have emerged over the years. It is noted in passing that views on what type of technique CBA really is, or should be, have differed and continue to differ depending on the set of value judgements used. We emphasise that the more recent suggested extensions to or modifications of CBA have sought to make the technique more comprehensive (i.e. to include distributional, environmental quality and other objectives as well as economic efficiency) at the inevitable cost of a loss of precision. The underlying principles of CBA are examined and the problems of multiple objective planning highlighted. Valuation problems are analysed, in particular with regard to amenity and environmental effects. Finally, the issues of uncertainty, irreversibility and intergenerational equity are raised briefly to indicate the complexity of the decision-making task when large-scale technologically advanced projects have to be appraised. In conclusion it is argued that CBA presented in a disaggregated format, as a comprehensive method for the ordering of information and a testing procedure for a range of valuations can perform a useful role in the decision-making process. The term 'policy analysis' is probably a better one for this broader view of the use of CBA. Ultimately we seem to be searching for a synthesis of the participatory and the traditional (technocratic) style of decision-making though we have a long and difficult course to chart before we even approach such a goal.

INTRODUCTION

THE MAIN purpose of this paper will be to survey some of the theoretical and practical criticisms that have been made of formal cost–benefit analysis (CBA); and also to try and indicate the different views concerning the most 'useful' role (in terms of goal fulfilment) that the formal technique and its extensions are capable of performing in the decision-making process. We shall first outline the main limitations of the formal technique and indicate other possible more comprehensive but less precise alternative or complementary techniques such as policy analysis and environmental impact analysis (EIA) in order to draw some conclusions about the role that CBA could most usefully fulfil.

CBA can be defined as a formal procedure for resource allocation in which all the relevant costs and benefits of a public investment decision or course of action are analysed. In practice the technique has been used to enumerate and evaluate a wide range of projects such as river management schemes, motorway and airport construction, health and education services and many others. The actual extent of government intervention and investment in the economy has, of course, varied from country to country and the precise nature of the CBA used has also varied both between government departments and institutions in the same country as well as internationally. Formal CBA seeks in principle to establish rules by which the decision-maker can identify efficient projects (i.e. those leading to a net increase in national income). A project yielding positive discounted net benefits in present value terms

passes the test [32]. The concentration of formal CBA on the efficiency objective has meant giving only secondary attention to other possible and no less legitimate objectives such as distributional equity, environmental quality or national security. Formal CBA attempts to mimic the decisions that a perfectly competitive (Pareto efficient) market mechanism would have arrived at and apply them in the public sector.

Despite an earlier heritage the widespread application of CBA is usually dated from the implementation of the US water project planning programmes in the 1930's. Dorfman [9] points out that over the years, with increasing government involvement in the economy, CBA has been applied to an ever wider spectrum of projects. In its early applications the technique was used to evaluate projects which basically produced intermediate outputs—the classic examples being irrigation projects designed to stimulate farm productivity. The essential point to note is that such projects yielded outputs which were generally amenable to evaluation via market prices. As the project spectrum has widened to cover the fields of education, transport, health, recreation and other environmental services provision, however, CBA has increasingly appeared to many critics to be less useful than at first thought. For some critics the technique's early promise has flattered only to deceive in the face of projects, for example, producing outputs of an intangible or regionally localised nature. Difficult problems of evaluation in the absence of adequate market prices (shadow pricing) and of distributional equity have been recognised and have served to increase the growing doubts voiced by some commentators concerning the precise value of CBA as a decision-making tool [1, 46, 12]. Herfindahl and Kneese [19] note that governments have increasingly sought evaluation of projects involving small initial investments but requiring operation, maintenance and replacement cost streams stretching out into the distant future; while other projects have required 'implicit' rather than direct investment. Thus the preservation of the UK's Vale of Belvoir or a North American 'wilderness area' in their 'natural' states involves the loss of the explicit benefit streams that would be produced by mining or forestry operations. On the other hand such projects if given the go ahead would in many cases involve irreversible effects and the possible consequent foreclosing of options, not to mention the thorny problems of intergenerational equity and decision-making with uncertain costs and benefits [37, 23].

Over time the projects that CBA could have been confronted with have been increasing both in size and complexity. Rational decisions have been required, for example, with regard to nuclear power generation programmes and waste recycling plants; large resource extraction programmes such as the proposed Vale of Belvoir and Selby Coalfield developments (the former involving 'dirty' seam mining and consequent spoil disposal problems); North Sea oil developments and flood protection schemes such as the Thames Barrage and the proposed Yare barrier in Norfolk. Inevitably such schemes produce complex environmental and distributional effects (both intragenerational and intergenerational); and a number stimulate difficult 'social disruption' effects because of their rural location and consequent disruption to small often fragile local economies and communities [28, 5]. A growing general awareness of the need for some measure of environmental protection and consideration of environmental values has led to the construction of so-called EIAs at some stage in the planning process for a number of projects in the USA and a call for similar, though not identical, analyses to be implemented in the UK [7, 8, 52].

We indicated earlier that formal CBA operates on the basis of an economic efficiency criterion with the objective of maximising national income to the neglect, in practice, some critics say, of other possible but less quantifiable objectives by which public investment projects could be judged. Freeman [14] notes that a single government expenditure objective, the improvement of human welfare, would probably be dominant in an open and rational democratic society. However, since we have no direct measure of welfare this ultimate goal can give little practical guidance to decision-makers. Steiner [48] argues that given multiple planning objectives individual projects are bound to affect more than one dimension and the objectives themselves will often conflict. The central problems of multiple objective planning become how to define an appropriate

Vol. 7
1979

measure for each objective and how to resolve the problem of conflicting objectives. It is to these problems that we now turn our attention.

THE UNDERLYING PRINCIPLES OF CBA

The terms benefits and costs can only be defined precisely with reference to a particular decision criteria and underlying objective. Nash *et al.* [35] argue further that any decision criterion must reflect some set of value judgements which determine what effects should be regarded as benefits and costs; so that there can be no uniquely 'proper' way to do CBA. The welfare economics sub-discipline has offered some criteria which have tempted some people to possibly regard them as the 'correct' basis for decision-taking [18]. The Pareto efficiency criterion, for instance, lays down that any project or course of action which makes some individuals in society better off without making anyone worse off should be undertaken. This criterion suffers two severe disadvantages. It is highly restrictive because real projects and courses of action nearly always involve losses to some individuals and benefits to others and thus do not fit the Pareto rule. In the second place the criterion is extremely conservative in matters of income distribution, essentially preserving the prevailing distribution in society by means of an implicit equity weighting which favours the *status quo*.

Two more general criteria, the Hicks–Kaldor compensation criterion (termed a 'potential Pareto improvement' criterion by Mishan [32]) and Little's distribution criterion [25] were developed in order to try and overcome the disadvantages of the Pareto rule. The compensation principle lays down that any project which yields sufficient benefits to some individuals (measured by willingness to pay) such that these gainers could compensate the losers and still remain better off themselves is desirable. Mishan [32] has argued that cost–benefit analysts should confine themselves to procedures that do in fact produce a potential Pareto improvement in welfare. Both the criteria outlined so far entail the acceptance of the value judgements that individual preferences should count and that these preferences are weighted by market power. Hypothetical compensation could well involve inequitable redistributions of income in society. Little [25] reformulated the basic compensation principle suggesting that a project be accepted if the sum of its discounted net benefits is positive *and* if it does not cause a deterioration in the prevailing distribution of income. This statement, however, only serves to highlight the practical operational problems of how a distributional weighting system for costs and benefits is to be constructed and trade-off ratios between efficiency and equity established.

Some analysts seem to argue that the economists role should be to advise government on efficiency matters alone and to let the government decide, through direct instruments, the question of distributional equity [34]. Steiner [48] is doubtful, however, about complete reliance on direct methods to correct for overall maldistributions of income. He argues that social change is such that this would in practice overweight the current distribution of income and would constrain any redistribution. A number of approaches seeking to integrate efficiency and equity effects can be identified in the economics literature. One such approach involves the application of weights to different arguments in the objective function in order to derive a single-valued measure of benefits. Within this general approach there are two viewpoints concerning the weighting procedure. One view seeks to establish weights prior to the taking of any decisions either by direct consultation with decision-makers or by detailed analysis of previous government decisions on resource allocation and taxation [11, 50, 16, 29]. Other researchers take the view that weights can only be established by an iterative political decision-making process working on a case by case basis [26, 30]. Reliance is therefore placed on the choices actually made by policymakers. Discovering what any government's preferences really are is a formidable task and none of the suggested weighting systems are without limitations. Nash *et al.* [35] conclude that advocates of what they term this 'Management Science' approach will need to use sensitivity analysis. A series of welfare-weighted calculations using alternative weighting functions should be presented to the policymaker.

An alternative approach to the problem of objective integration rejects the notion of a single-valued measure of benefits because of

the severe difficulties of ensuring that a range of public policy objectives can be made commensurable. Steiner [48] believes that the single number measure of benefits approach in practice submerges real issues behind a facade of faulty measurements. Williams [51] also sees little scope for systematic incorporation of redistributive effects in CBA but stresses the need for incidence effects to be included as separate exhibits in the analysis presentation. Lichfield [24] too stresses the importance of presenting CBA in a disaggregated fashion. The final judgement on conflicting objectives is left to the policymaker and no attempt is made to articulate how this judgement should be made. Other economists [19] appear to support much of Steiner's argument but see the calculation of distributional effects as essentially supplementary to the dominant efficiency calculations.

VALUATION PROBLEMS

We noted earlier that CBA has been applied to an ever wider range of projects and that both the scale and complexity of many of these projects has increased dramatically. Complex problems have therefore been set for the analyst attempting to attach money prices to costs and benefits in order to reflect society's valuations of the goods and resources involved. Despite the optimism of some economists the practical difficulties presented by, in particular, environmental costs and benefits are immense and the current state of the art cannot be said to represent anything like an adequate solution to the social cost benefit valuation problem [39, 27]. McKean [31] is not optimistic that shadow prices will ever be accurately constructed for a range of intangible effects and stresses the high informational costs involved in even a small increase in accuracy. Williams [51] shares this fairly pessimistic view but is prepared to accept that there is a role in CBA for the 'postulated' price. He defends its use on the grounds that it at least provides a focus for discussion by others who might disagree with the policymaker's judgement.

A number of ingenious valuation techniques have been developed in order to produce proxy or surrogate values for the unknown true shadow prices. These proxy values, however, can at best be regarded only as lower limits to the 'true' values of various environmental costs and benefits. The techniques themselves include survey or questionnaire methods to gauge willingness to pay or compensation. There are two basic obstacles to the successful implementation of such methods: the problem of interviewee perception (individuals often find it difficult to perceive the personal implications of hypothetical changes in environmental quality); and the problem of incentive and response distortion where individuals may not reveal their true preferences if, for instance, the question of taxation is involved. Other techniques operate on the principle that it may be possible to deduce surrogate monetary amenity losses or benefits from observation of consumer behaviour. Thus the Clawson–Knetsch method [21] attempts to estimate recreational benefits through observation of recreationists and the amount of money they are willing to spend on travel costs to a site. This method has been heavily criticised over the years in the literature [13]. Methods utilising the so-called 'hedonic' pricing approach assume that changes in and patterns of property and land values will reflect external amenity or disamenity again allowing a proxy measure of the shadow price to be derived. Studies using this approach have also been severly criticised [40].

Bradford and Feiveson [4] argue that the inability of analysts to construct monetary surrogates for important intangible values has had little perceptible effect on the degree to which they will be regarded or disregarded by policymakers. The authors are generalising on US experience and they do insert the caveat that often efforts were made to inform the policymakers about the intangible effects by detailed presentations of physical data. Other factors which will also determine whether intangible factors get incorporated into the decision-making process will include the political importance of the issues, the degree to which certain values have general rather than minority support and the preferences and values of the policymakers themselves. Clearly it is difficult to generalise internationally but it is not difficult to think of UK cases where intangible effects have appeared to get very little consideration.

We will now examine briefly one illustrative example, the proposed Yare barrier flood protection scheme in Norfolk, without wishing to claim too much in the way of generality. The

Vol. 7
1979

official CBA report [43] had in its terms of reference "the need for the preservation of the existing character of Broadland and the conservation of the area's unique ecological habitats". Despite this the report devotes just a couple of pages to the possible ecological effects (judging them to be insignificant) and neglects completely any amenity issues. If constructed, however, the barrier is expected to stimulate a tenfold increase in arable farming acreage. Indeed the project meets the criterion of positive discounted net benefits largely due to the estimated agricultural benefits stimulated by an expansion in wheat cropping. This change in farming practice will mean the complete disappearance of portions of the characteristic Broadland valley grazing marsh landscape. Some of this area could reasonably be described as of critical landscape significance and includes marsh dykes which are still rich in flora. There will be keenly felt amenity and aesthetic losses for some groups, though perhaps not for all groups (it is claimed that for some water recreationists it does not really matter whether they pass by marshland or arable fields) as well as ecological losses. The extent of such losses must nevertheless be qualified by the fact that agricultural practice has changed in the past and will continue to change in the future whether the barrier is built or not. Paradoxically though, the greater the rate of this 'ongoing' change the more difficult it will be to gauge the estimated agricultural benefits supposedly dependent on barrier construction. Both the sample size and questionnaire methods used in the CBA to determine rates of arablisation with and without the barrier are open to criticism [47].

Andrews and Waits [2] take the viewpoint that the whole debate over whether or not environmental values can or should be quantified is not a meaningful one. They argue all such values can be quantified either ordinally or numerically or monetarily. The important question for them is what indicators provide appropriate representations of environmental values, and how precise they are. They call for a further comprehensive research effort to answer this question. Brooks [6] believes that it is possible to formulate fairly precisely, if not quantitatively, some implicit criteria for the valuation of nature. He lists the concepts of uniqueness, reversibility or resilence, aesthetic beauty, human life support capacity and 'naturalness'. Brooks agrees that there is no independent, rational basis for assigning a price to such criteria in a CBA, but argues that criteria definition is an important first step that aids analysis by permitting testing of the impact of various assigned prices.

CONSERVATION, IRREVERSIBILITY AND INTERGENERATIONAL EQUITY

The problems of quantifying intangible effects can lead one to the question of what role the human species ought to play in the natural order of things. Leaving on one side the possible ethical-religious criticisms of man-centered CBA a more moderate 'Environmentalist' position would be that society should utilise CBA as a decision making tool subject to the constraint "that a broad range of animal and plant life exists in a substantial geographic range under conditions these life forms find congenial" (ecological diversity) [1]. Given limited funds a trade-off must somehow be struck between preservation and the satisfaction of more pressing human wants. It would seem reasonable to first try and protect all of the most 'valuable' of the landscapes and ecologically important or fragile areas, assuming that such areas can be identified using perhaps the Brooks' criteria. Krutilla and Fisher [22] remind us that the natural biota represent our reservoir of genetic information, which has economic value. Advances in agriculture in terms of improved crop varieties can depend on genetic information found in the diversified natural biota. The same biota has in the past also provided the basis for a wide range of medicinal drugs and only a small portion of the potential medicinal value of biological specimens has yet been tapped.

In circumstances where environmental effects are uncertain and irreversible, as is the case for many of the larger projects, the quantification and valuation problems become exceedingly complex. Following Hirschleifer and Shapiro [20] the terms 'risk' and 'uncertainty' express a situation in which analysis requires us to take into account the possibility of a number of alternative outcomes or consequences of actions. Risk aversion is then interpreted to mean that the very existence of un-

certainty is regarded by individuals as a cost. On the basis of the so-called Arrow–Lind Theorem [3], however, it has been argued that there are circumstances when individuals may be risk averse but society as a whole may be risk neutral; and that, therefore, uncertainty can be neglected in social evaluations. The underlying idea here rests upon the law of large numbers in statistical theory. A number of writers [17, 23] have suggested, nevertheless, that environmental damages, especially irreversible damages, are probably sufficiently different to prevent them being subject to the above theorem. Irreversible environmental damages are significant effects which permanently, or for a very long period of time, reduce the variety of future choices. Thus projects such as large scale strip mining, wetland drainage and arablisation or dam construction with its consequent valley flooding all come under the umbrella of irreversibility; as does pollution from stock pollutants such as mercury, cadmium or radioactive waste.

Irreversible damage effects may serve to reduce individuals' options to experience the environmental good or service in question i.e. hike or picknick in the previously dry valley behind the dam or on the hilltops levelled by mining activities. An individual's option value [36] can be defined as a benefit accruing to an individual as a result of his retaining an option to consume the good at some time in the future. In effect it is a premium for risk avoidance, an additional willingness to pay to insure oneself against the irretrievable loss of some environmental good. Irreversibility is not, however, an easy concept to pin down and a moment's thought leads one into questions of absolute irreversibility or long term irreversibility with the possibility of eventual retrieval (this latter possibility being dependent on costs and technological advance). Further, there is the debate over the 'authenticity' of the restoration process; clearly the outdoor recreation market is in reality a complicated structure of numerous submarkets (we touched upon this question earlier when discussing the boating recreationalists' preferences on the Norfolk Broads).

The difficulties inherent in dealing with the issues of risk, uncertainty and irreversibility are very great and the current state of the art in economics is a long way from providing full solutions. Practical CBA typically employs the assumption of a certain world or incorporates some rather imprecise (but perhaps not dangerously incorrect) adjustments to take account of uncertainty. Haveman [17] suggests, as far as uncertainty in the costs stream is concerned, the following adjustments:

(a) increase the stream of costs according to the degree of risk aversion which the policy-maker feels is present;

(b) lengthen the time horizon over which the damages are expected;

(c) reduce the discount rate used in the present value calculations.

Krutilla and Fisher [23] conclude that a conservative decision-taking policy is required with respect to irreversible environmental effects. They argue that if the costs and benefits of alternative uses of the environment are not known with any certainty then for a risk averse society there will be a value in retaining an option to use the environment in a way that could otherwise be foreclosed. It is necessary to bear in mind, however, (as the authors do) that this path breaking work has not yet faced squarely the important question of whether or not any particular real course of action is in fact irreversible (the analysis assumes irreversibility to be a fact). Thus we have not yet reached the level of generality in this complex area necessary for real world decision making. Suffice it to note at this stage that a policy of non foreclosure of options may in practice turn out to be very restrictive.

The problem of irreversibility is intimately bound up with the problems of intergenerational social choice criteria and the concept of intergenerational equity [41]. Page [37] demonstrates that questions of resource depletion and stock pollutants are fundamentally questions of equitable distribution of burdens across generations. For Page the equity question is, what is a fair distribution of risk to impose upon the future? He concludes that there is no easy way to add up the costs and risks along with the benefits and no way to guarantee that future generations will enjoy a higher level of welfare than the present. The discounting method used as an intergenera-

Vol. 7
1979

tional social decision rule can ensure intergenerational efficiency but in terms of equity it only serves to bias choice in favour of the present generation [38]. When the losers from a project cannot vote as is the case for future generations, the hypothetical Pareto improvement criterion breaks down. Freeman [15] suggests that compensation to future generations (the losers) actually be paid by setting aside a sum of money now which when cumulated forwards at the discount rate (now a familiar rate of interest) would provide monetary compensation for future losers. Page [37] adopts a Rawlsian standpoint [42] on the question of intergenerational equity. The view carries with it the presumption of equal distribution with bias towards the least well-off. For Page the least well-off are likely to be future generations whose welfare the current generation is reducing by favouring policies which over-exploit resource bases, lead to accumulations of stock pollutants in the biosphere and destroy unique landscapes and ecosystems. Mishan [33] offers a dissenting view to those described above and concludes that the economist has really nothing of substance to contribute to the debate on the 'ideal' intergenerational time path.

THE ROLE OF CBA

Rowen [44] identifies two different points of view towards CBA. One, which he argues has often been expressed in the economic literature on CBA, sees the objective of CBA to be to make a choice. A decision rule is constructed which allows efficient choices to be made. To what extent this view is still generally held by economists is, however, open to debate, a number of writers being difficult to pin down on this score. There is less doubt that the economists' work has been used by others either to justify past choice or as a delaying tactic. Dorfman [10] notes that CBA has suffered from being an economic approach to a political problem. The results of CBA were wrongly presumed to be determinative and the technique was unrealistically expected to fulful a role it was not capable of discharging. Rowen argues that the decision-rule view of CBA is of limited utility because over a range of policy areas and decisions, individuals (including decision-makers) do not have preferences among broad goals, nor, in general, among overlapping policy objectives.

The second view of CBA identified is one which sees the technique as a means of assisting choice. It is recognised that CBA does not eliminate the need for the political resolution of conflicting objectives but that it can assist in the process of objective formulation and choice. The real debate here is how far CBA can go in assisting choice. Mishan [32] cautions economists undertaking CBA to utilise the potential Pareto improvement criterion and no other. This would seem to many to be an overly restrictive role for CBA to play. But to play a broader role traditional (Paretian) CBA must be extended to encompass not just efficiency but many other objectives as well. We have seen that some analysts see multiple objective planning as the search for a single-valued benefit measure while others reject this approach on practical grounds and leave the final judgement to the policymaker.

CBA interpreted in a broader context (policy analysis) can be of great utility to the policymakers identified by Rowen as lacking clearly articulated or well defined preferences. Policy analysis can make an essentially heuristic contribution in this situation providing a conceptual framework for relating resources to objectives, an ordering device for information and exposing the individuals to a range of objectives. Many commentators have stressed the need for CBA or policy analysis, if it is to be a genuine component of democratic policymaking, to be presented in a disaggregated format. The aim must be to encourage a wider public participation in decision-making and this can only be achieved by presenting information in a more accessible and understandable manner. Tribe [49] argues that the process by which a decision is arived at is often at least as important as the result itself. There is a danger, nevertheless, that participatory decision-making may involve such time delays that projects (even desirable and much needed ones) will be effectively hamstrung. For Brooks [6] the ultimate difficulty with participatory decision processes is the lack of assurance that all the relevant interests and perspectives will be represented in a balanced way.

Perhaps a new synthesis of the participatory and the old technocratic style of decision-making is required. Dorfman [10] argues that a

number of what he terms adversary analyses ought to be presented in any major project decision problem, and the funding of such analyses should be controlled by some quasi-judicial body. Brooks [6] would add a second component, a technical–analytical court, that is capable of reviewing the various analyses presented and which would present an understandable overview of the choices for public and policymaker consumption. The final judgement would still, however, rest with the policymakers. The establishment of such a 'neutral' court would seem to be an enormously difficult if not impossible task.

REFERENCES

1. ACKERMAN B *et al.* (1974) *The Uncertain Search for Environmental Quality.* The Free Press, New York, USA.
2. ANDREWS R & WAITS M (1978) *Environmental values in public decision.* School of Natural Resources Report, University of Michigan, Ann Arbor, USA.
3. ARROW K & LIND R (1970) Uncertainty and the evaluation of public investment decisions. *Am. Econ. Rev.* **60**(3), 364–378.
4. BRADFORD D & FEIVESON H (1976) Benefits and costs, winners and losers. In *Boundaries of Analysis: An Inquiry into the Tocks Island Dam Controversy* (Eds FEIVESON H *et al.*) 125–161, Ballinger, Cambridge, Massachusetts, USA.
5. BROADY M (1978) Method and methodology in social impact analysis. Paper presented to The Institution of Water Engineers and Scientists Symposium on Engineering and the Environment: *Harmony or Conflict?* London, UK.
6. BROOKS H. (1976) Environmental decision making: analysis and values. In *When Values Conflict* (Eds TRIBE L *et al.*) 115–135, Ballinger, Cambridge, Massachusetts, USA.
7. CATLOW J. & THIRLWALL C (1977) *Environmental impact analysis.* DoE Res. Rep. **11**, Department of the Environment, London, UK.
8. CLARK B *et al.* (1978) Methods of environmental impact analysis. *Built. Env.* **4**(2), 111–121.
9. DORFMAN R (1976) *Forty years of cost benefit analysis.* Harvard Institute of Economic Research, Discussion Paper **493**, Harvard University, USA.
10. DORFMAN R (1976) An afterword: human values and environmental decisions. In *When Values Conflict* (Eds TRIBE L *et al.*) 153–173, Ballinger, Cambridge, Massachusetts, USA.
11. ECKSTEIN O (1961) *A Survey of the Theory of Public Expenditure Criteria, in Public Finances: Needs, Sources and Utilisation.* Princeton University Press, Princeton, New Jersey, USA.
12. FEIVESON H *et al.* (1976) *Boundaries of Analysis: An Inquiry into the Tocks Island Dam Controversy.* Ballinger, Cambridge, Massachusetts, USA.
13. FLEGG A (1976) Methodological problems in estimating recreational demand functions and evaluating recreational benefits. *Jl Reg. Stud.* **10**(3), 353–362.
14. FREEMAN M (1977) Project Design & Evaluation with Multiple Objectives. In *Public Expenditure and Policy Analysis* (Eds HAVEMAN R & MARGOLIS J) 239–256, Rand McNally, Chicago, USA.
15. FREEMAN M (1977) Equity, efficiency, and discounting. *Futures.* **9**(5), 375–376.
16. HAVEMAN R & WEISBROD B (1977) Defining benefits of public programmes: some guidance for policy analysts. In *Public Expenditure and Policy Analysis* (Eds HAVEMAN R & MARGOLIS J) 135–160, Rand McNally, Chicago, USA.
17. HAVEMAN R (1977) The economic evaluation of long-run uncertainties. *Futures.* **9**(5), 365–374.
18. HEAD J (1974) *Public Goods and Public Welfare.* Duke University Press, Durham, North Carolina, USA.
19. HERFINDAHL O & KNEESE A (1973) *Economic Theory of Natural Resources, Resources For the Future.* Charles E Merrill, Ohio, USA.
20. HIRSCHLEIFER J & SHAPIRO L (1977) The treatment of risk and uncertainty. In *Public Expenditure and Policy Analysis* (Eds HAVEMAN R & MARGOLIS J) 180–203, Rand McNally, Chicago, USA.
21. KNETSCH J (1974) *Outdoor Recreation and Water Resources Planning.* American Geophysical Union, Washington, USA.
22. KRUTILLA J & FISHER A (1975) *The Economics of Natural Environments. Resources For The Future,* Johns Hopkins Press, Baltimore, USA.
23. KRUTILLA J & FISHER A (1975) Valuing long run ecological consequences and irreversibilities. In *Cost Benefit Analysis and Water Pollution Policy* (Eds PESKIN H & SESKIN E) 271–290, The Urban Institute, Washington, USA.
24. LICHFIELD N (1968) Economics in town planning. *Tn Plann. Rev.* **39**(1), 5–20.
25. LITTLE I (1957) *A Critique of Welfare Ecomomics.* Oxford University Press, Oxford, UK.
26. MASS A (1966) Benefit–cost analysis: its relevancy to public investment decisions. *Q. Jl Econ.* **78**(2), 208–226.
27. MALER G & WZGA R (1976) Economic Measurement of Environmental Damage, OECD, Paris, France.
28. MCEVOY J & DIETZ T (1977) *Handbook For Environmental Planning* John Wiley, New York, USA.
29. MCGUIRE M & GARN H (1969) The integration of equity and efficiency criteria in public project selection. *Econ. Jl.* **79**(316), 882–893.
30. MCKEAN R (1958) *Efficiency in Government Through Systems Analysis* John Wiley, New York, USA.
31. MCKEAN R (1968) The use of shadow prices. In *Problems in Public Expenditure Analysis* (Ed CHASE S) 33–77, Brookings Institution, Washington, USA.
32. MISHAN E (1975) *Cost Benefit Analysis.* Allen & Unwin, London, UK.
33. MISHAN E (1977) Economic criteria for intergenerational comparisons. *Futures.* **9**(5), 383–403.
34. MUSGRAVE R (1969) Cost benefit analysis and the theory of public finance. *J Econ. Lit* **7**(3), 797–806.
35. NASH C *et al.* (1975) An evaluation of cost-benefit analysis criteria. *Scott. Jl Polit. Econ.* **22**(1), 121–134.
36. OLSEN G (1975) Option value. *Aust. Jl Agric. Econ.* **19**(3), 197–209.
37. PAGE T (1977) *Conservation and Efficiency, Resources for the Future.* Johns Hopkins Press, London, UK.
38. PAGE T (1977) Discounting and intergenerational equity. *Futures.* **9**(5), 377–382.
39. PEARCE D (1978) *The Valuation of Social Cost.* Allen & Unwin, London, UK.
40. PEARCE D & EDWARDS R (1979) The monetary evaluation of noise nuisance. In *Environmental Planning and Resource Management,* (Eds O'RIORDAN T & D'ARGE R) **1** (1), John Wiley, London, UK.

Vol. 7
1979

41. PEARCE D (1977) Accounting for the future. *Futures.* **9**(5), 360–364.
42. RAWLS J (1974) *A Theory of Justice.* Oxford University Press, London, UK.
43. RENDELL, PALMER & TRITTON (1978) *The Yare Basin Flood Control Study*, **1–3**.
44. ROWEN H (1975) The role of cost–benefit analysis in policy making. In *Cost Benefit Analysis and Water Pollution* (Eds PESKIN H & SESKIN E) 361–370, Urban Institute, Washington, USA.
45. ROWEN H (1976) Policy analysis as heuristic aid: the design of means, ends, and institutions. In *When Values Conflict* (Ed TRIBE L), Ballinger, Cambridge, Massachusetts, USA.
46. SELF P (1975) *Econocrats And The Policy Process.* Macmillan, London, UK.
47. SCHOOL OF ENVIRONMENTAL SCIENCES (1979) *The Yare Basin Flood Control Study: Some Critical Comments.* University of East Anglia, Norwich (mimeo).
48. STEINER P (1977) The public sector and public interest. *In Public Expenditure and Policy Analysis* (Eds HAVEMAN R & MARGOLIS J) 27–66, Rand McNally, Chicago, USA.
49. TRIBE L (1976) *When Values Conflict.* Ch. **3**, Ballinger, Cambridge, Massachusetts, USA.
50. WEISBROD B (1968) Income redistribution effects and benefit–cost analysis. In *Problems in Public Expenditure Analysis* (Ed CHASE S) 177–222, Brookings Institution, Washington, USA.
51. WILLIAMS A (1972) Cost–benefit analysis: bastard science? and/or insidious poison in the body politick? *Jl Publ. Econ.* **1**(2), 199–226.
52. WILLIAMS K *et al.* (1978) EIA and the vale of Belvoir coalfield. *Built Env.* **4**(2), 142–151.

ADDRESS FOR CORRESPONDENCE: *RK Turner, Esq, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK.*

BLANK
PAGE

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 5, pp. 421 to 430
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1001-0421$02.00/0

Economic Approaches to Environmental Decision-Making

J LOWE

University of Bath, UK

D LEWIS

University of York, UK

(*Received May* 1979)

There are several important economic reasons why government has a role in environmental management. However there are also substantial informational problems that must be solved if government intervention is to be adequately directed. Traditional cost–benefit techniques are subject to some serious conceptual and empirical problems. However there are less powerful approaches which provide more feasible alternatives to environmental measurement which the authors discuss.

INTRODUCTION

THE CONTROL of the environment presents considerable managerial problems for those responsible for this facet of public decision making. These problems stem from several basic causes. In common with many parallel cases in the public sector, the environment is not a product that is easily traded between interested parties, consequently there are few price signals which help guide the decision makers courses of action. This may be partly due to inadequate information and it is information or lack of it which causes several general problems for the decision maker. In particular, environmental decisions have to be made under circumstances when the current state of knowledge is incomplete. Complete information on the side effects and environmental impact of various chemicals etc. rarely precedes their use and for example the effects of DDT and mercury have been grossly miscalculated in the past. Recently there has been substantial controversy over the impact of lead on the mental development of children. Within the space of a few months two reports from highly reputable bodies gave substantially conflicting views on the effects of lead additives in petrol [1, 2].

Even when the decision maker is aware of the environmental consequences of particular actions, he has to make value judgements on a number of factors. In particular many forms of environmental degradation, either directly or indirectly, may be associated with shortening of human life expectation. This clearly causes enormous problems of valuation and judgement. Also trade-offs have to be made between environmental harm and the reduced economic growth, prosperity and employment that may result from devoting resources to pollution control. Most importantly the environmental decision maker has to make judgements about the distribution of economic and environmental goods between current and future generations—even though periodic changes of government resulting in ballot box appeals to the current generations may make this task dif-

ficult. Finally the organisational structure of the controlling bodies may themselves lead to problems of environmental management. In particular when a system exists (as in the UK) where separate bodies are responsible for the air, water, and land, then the likelihood of non-optimal decisions from the point of view of the environment as a whole is increased as many forms of pollution control merely switch pollution from one media to another.

Some of the issues mentioned above may be relatively intractable whilst others can be solved by a change in control agency knowledge and approach. The aim of this article is to evaluate how the public decision maker can utilise and change his information system in order to improve environmental decisions. In particular the paper focuses on some of the empirical difficulties that may be encountered and their solution.

ENVIRONMENTAL DECISIONS: A VOLUNTARY APPROACH

The basic problem caused by environmental pollution is that it is frequently generated by an economic unit that does not itself bear all the consequences of its actions. It was Ayres and Kneese [3] who pointed out that external costs and benefits are liable to occur in almost any productive process in the form of residuals whose disposal is not costless. Indeed the final product of an industry will in general become a residual itself eventually, since consumption typically involves the enjoyment of a flow of services provided by the good rather than the good's physical destruction. Of course, not all residuals inherently impose a cost on society and many can be converted into inputs for some other process by the use of further resources, with the result that the external effect of a residual flow can be positive, negative or zero. The important point is that, given the very widespread occurrence of substantial environmental externalities in production activities, which result in a divergence between private and social cost, there is a need for government action to internalise the externality to ensure an optimum allocation of resources.

However there have, inevitably, been suggestions that this optimum level of production can be achieved without necessitating any public intervention. In particular two voluntary solutions often put forward are that the polluter is socially responsible and controls his own environmental degradation or alternatively that those impacted by the pollution, and those responsible for it, come to some negotiated agreement about pollution output levels.

Whilst many firms do engage in programmes of corporate social responsibility, experience suggests that this is rarely enough to deal with problems of the environment. The second solution deserves greater consideration since it has been discussed extensively in the literature and deals with some important philosophical issues. Ronald Coase was its first proponent [4] but since then many other economists have concerned themselves with the so called 'bargaining' or 'voluntary' solution. An excellent survey of this work is by Dick [5].

The voluntary approach to the externality problem revolves around two basic principles. Firstly it is suggested that since in cases of environmental polution there are gainers and losers, then it should be possible for these two parties to bargain in such a way that the losers can bribe the gainers to reduce pollution levels. Consequently an allocation of economic resources, the same as state intervention would seek, could be achieved as the losers from pollution could bribe the gainers up to the point where marginal private benefit of production is equal to its marginal social cost. Secondly it is suggested that the outcome of such a bargaining process is invariant with respect to the legal ownership rights of the environmental resources that are used both for life support and as a waste sink.

Whilst these arguments have considerable intuitive appeal it is doubtful whether they are either internally consistent or could serve as the basis of policy towards the environment. A major facet of the approach is that it largely assumes away the transaction costs and organisational problems of getting those impacted by the pollution to act as a group. Even if it were assumed that they could act in unison then the negotiated agreement becomes a bilateral monopoly case which *a priori* is likely to lead to an indeterminate solution. Finally the possibility of threats could also disturb any cosy notions of equilibrium. It is almost axiomatic that good bargainers rarely disclose their

Vol. 7
1979

bargaining limits—consequently threats become a likely aspect of negotiation. In the presence of threats any agreement is only likely to be optimal by chance. Indeed threats may induce polluting behaviour beyond that level profitable for the polluter in order that his threats become more believable in the next negotiating round.

The independence of ownership rights and the level of activity also seems a dubious proposition both in theory and practice. In the long run the bribe revenues collected by the polluter (if the ownership rights had been ascribed to him) would act as an inducement for new firms to enter the industry and thus change production levels and the allocation of resources. Finally, a compelling reason to reject the negotiated solution is that, in view of the long run existence of serious externality problems that demand some form of government intervention, the voluntary approach has clearly not been workable.

A final decentralised alternative would be for government to change the legal status of all environmental resources such that those valuing the environment for consumption and those valuing it for purposes of waste disposal could in theory bid for the rights to use parts of various environmental media. Thus at each stage in the bidding process the bids would represent incremental benefits that various parties attach to environmental resources. In theory the concept is fine—especially if the assimilative capacity of the environment—and thus the welfare of future generations was protected by a minimum environmental standards which could not be the subject of auctioning. However such a scheme would still run into problems of individuals not participating in the bids but still receiving the benefits that were due to those who did. This is the problem of public goods provision which is a central reason for state intervention. Also if effective bidding coalitions were to be formed it is likely that there would be substantial transaction costs and organisational problems in their formation.

It seems reasonable to conclude, therefore, that because of the essentially public nature of environmental benefits neither voluntarily negotiated solutions not the creation of private markets represent a reasonable approach to the management of environmental resources.

COST–BENEFIT EVALUATION AND THE ROLE OF GOVERNMENT

In the absence of a market solution to environmental problems, public decision makers must decide between alternative courses of action on the basis of a cost–benefit evaluation. Following this they may control environmental degradation by standards, taxes, or subsidies. In this paper however we do not concern ourselves with the mechanism of controls but instead concentrate on the information requirements for effective control decisions.

The underlying philosophy of cost–benefit analysis is that all the costs and benefits of a particular action should be considered, to whosoever in society they accrue, so that it would seem that this technique is appropriate for the appraisal of environmental policies involving the regulation of external effects on different groups of people. We concentrate below on determining the information requirements for effective decision-making, and discuss the viability of the use of cost–benefit analysis and an alternative approach in this context.

Clearly the frequency of incidence of substantial externalities involving the environment is a rationale for the use of social decision-making machinery. However the external cost of pollution to a particular environmental medium cannot be looked at in isolation. Abatement policies do not destroy matter, but only convert it from one form to another, an example being the incineration of solid waste causing air pollution. The problem is not how to dispose of residuals, but how to convert them into a form which imposes the least cost on society, including the cost of transformation. This cost must then be compared at the margin with the social benefit of the pollution-creating product.

Conceptual problems in the evaluation of environmental costs and benefits

It has been suggested there are severe obstacles which render the use of cost benefit analysis inappropriate for the analysis of certain kinds of environmental problems. Pearce [6] draws attention to the distinction between economic effects and biological effects. A good example of a form of pollution involving economic but not biological effects would be

glass bottles and the cost–benefit framework is capable of incorporating this kind of externality problem. The complex situation arises where biological forces are affected by an activity, but the economic effect is imperceptible.

Such a case would be the emission of cadmium, which above a certain level of concentration can have very serious health implications. In reality, of course, there is an economic effect associated with such an emission, but it may well go unperceived, due to the fact that damages emanate from the stock of pollution, rather than its flow. In the case of wastes which are essentially non-degradable, any addition to the stock of pollution imposes a damage cost for an indefinite period into the future, and this is difficult to ameliorate. Since the cost–benefit framework is designed to relate damage estimates to flows of pollution, it cannot easily be adapted for situations where it is the stock of waste that is the relevant concept, unless the damage from the stock pollutant is valued at infinity.

A separate, but somewhat related conceptual difficulty occurs where an excess discharge of waste impairs the assimilative capacity of the environment. This is referred to as a 'dynamic externality'. The problem here is that, given the assumption of instantaneous assimilation of waste which is within the environment's capacity, external costs only occur above this level. If there are net private benefits to production above this level, and the marginal external cost curve is upward sloping and continuous, then some level of waste emission above assimilative capacity must be deemed optimal.

If exceeding the environment's natural capacity impairs established degrader populations—and thus the assimilative capacity of the environment—then future generations will be able to discharge less and less to various environmental media and optimal production levels (where marginal external cost equals marginal private net benefit) will tend to fall towards zero. Clearly unless the external environmental cost function includes an element for the loss of assimilative capacity then equating costs and benefits as a guide to public decision making will inevitably lead to incorrect decisions. Ignoring discounting considerations this means that the optimum production level for society will be equal to that level which does not breach the assimilative capacity of the environment. Two caveats can however be made. Firstly, if society wishes to use a positive discount rate in its evaluation, then pollution levels in excess of what the environment can absorb will be justified. Secondly if it is conjectured that future new technology will enable a cleaning up of the environment to a greater extent, than is now possible, then clearly the use of a positive discount rate is wholly justified and breaches of the assimilative capacity can be allowed in the short run.

Alternatively it might be equally justifiable to assume a negative rate of discount for future cash flows. As time passes, and manufacturing processes throughout the world become increasingly extensive, environmental resources can be expected to become increasingly scarce and valuable. To treat air and water quality improvements in say, 50 years' time as virtually valueless, which is the impact of discounting, seems absurd. This is not to say that equal monetary benefits in different years should always be treated equally. It there are good reasons for valuing benefits accruing at a particular time more highly than others, then there is no theoretical objection to weighting these years individually. For example, if technological advance was expected to develop a pollution-free process for a particular industry in 10 years time, then the benefits of a treatment facility for the industry should be weighted to give low values for those accruing after this time. It is the 'straight line' treatment to flows of costs and benefits given in discounting, which is the subject of the valid objection.

It is probably true to assert that the reason discounting has been utilised so unquestioningly in environmental studies, is that without its use, cost–benefit analysis may sometimes lose its relevance. Many activities result in flows of costs or benefits lasting for an indefinite time period, which if summed without discounting have an infinite present value. This may render cost–benefit analysis superfluous to decision making. Conversely, the convenient use of an arbitrary discount rate results in the present value tending to a finite sum, allowing a neat comparison of cost and benefits. The fact that total environmental destruction in, say, 200 years' time would be minutely valued by this method, does not seem to matter. However despite the conceptual limitations cost–benefit analysis can usefully be employed for

Vol. 7
1979

examining many projects and strategies concerned with the environment. We now evaluate some of the practical valuation problems.

PROBLEMS IN THE VALUATION OF INDIVIDUAL COST AND BENEFIT COMPONENTS

There is a certain asymmetry in the ease of measurement of the cost and benefits of controlling pollution. The costs of pollution abatement are most easily measured at the level of the individual firm, since it is firms that keep the relevant data. Industry data necessarily present problems of aggregation which will be confused by the different age of plant, process differences etc. On the other hand, the benefits from the achievement of a particular level of reduction in emissions of a particular pollutant may be more readily estimated than the benefit from an emission reduction by an individual unit. This is because of the often complex relationship between each discharge and overall ambient conditions.

Costs

Essentially there are basically two types of pollution control; the first may be described as the short-run palliative which can often be added on to the production process either at the planning stage or retrospectively, frequently described as 'end-of-line pollution control'. This type of device includes, say, filtration plant used to clean exhaust gases before being discharged through a chimney. Whilst no change in the inputs (physical or capital) for the production process is involved, equipment is added to the end of the production line to control the discharge of pollutants to the atmosphere.

Generally, computing the cost to the firms of the various types of equipment is relatively straightforward. What cannot be costed, however, are the various 'intangibles' involved:— the 'cost' of lost production as the equipment is being installed, the opportunity of an alternative allocation of capital expenditure on productive plant and the opportunity cost of the physical space taken up by the control device. Quantification of this type of data is difficult but a subjective impression would be that in some cases they may be substantial and a simple totting up of the tangibles can seriously underestimate the full cost of pollution control. In addition small firms may lack the necessary expertise or have in-house resources to design control equipment. In these cases in particular the disruption costs of conforming to environmental regulation may be very substantial.

However, accepting the whole part of the end-of-line pollution control cost may also be misleading if the equipment brings benefits to the firm. Filtration plant, for example, can permit captured materials to be recycled and frequently in areas (such as non-ferrous metallurgy) where raw material costs are very high, this may be essential to ensure the economic viability of the process. The 'correct' cost to be allocated to pollution control is that involved for pollution control levels over and above those that would be undertaken solely for 'economic' reasons. Some recent research showed that few firms had any ideas about what proportion of their wastes they would collect for economic reasons. This may be partly due to the fact that pollution legislation pushed them into recycling when otherwise they may not have considered it [7].

The second type of pollution control measure is much longer-run in character involving a change in the nature of the inputs to the production process itself. It may mean the use of alternative raw materials processed in existing machinery or existing materials processed in new machinery or a combination of both. A wholly new technology may be introduced to manufacture the product. These changes can be classified as 'in-plant' and clearly these measures will frequently be incorporated over a very much longer period of time than the end-of-line control equipment. Thus, the length of time industry is given to clean up its emissions to the environment determines the nature of the control measures the firm is able to undertake and the extent and impact on its costs. We can identify many examples of process substitution which have had an impact upon pollution control; in some cases the move has been prompted by the need for pollution control but in others 'economic' and environmental considerations have worked in the same direction. An important example of a process change prompted almost entirely by pollution control considerations, is the now widespread adoption of the so-called 'pre-bake' cell in the production of aluminium [8]. This

is technically *less* efficient than other types of cell which have been developed but has the advantage that the waste gases are much more easily cleaned. In this case, pollution control necessitates the acceptance of a technically less efficient production process.

Various studies have now been made of the costs of environmental control and some of these are summarised by Lowe [9]. There are several aspects of these studies and the data generated which are important. Firstly there is a very high degree of variability even within very specific industries. In part this is because some of the most heavily impacted industries are those which have a mix of modern and old plant, which means the type of control and user needs will vary. Also locations are different and for a variety of economic, environmental, technical and political factors this may affect the costs of control and the degree of control achieved.

The size of firm and extent of integration also appear to be a determining factor of control costs since large and/or integrated operations which have substantial head office research resources are frequently better placed to negotiate and implement environmental control. This factor has been reinforced in some recent research which found for instance that the extent of recycling and firm size were closely linked in the Textile Dyeing and Finishing industry [10]. There is considerable evidence too of economies of scale existing in the operation of pollution control equipment. This, coupled with the technological problems faced by small firms, means that environmental control may force some firms to merge or go out of business altogether [11].

However, perhaps the most important aspect of costs is the marginal cost function. Here recent research [12] shows that control at levels much higher than at present, may in some industries lead to very substantial cost increases. One example quoted was in secondary lead processing where it was estimated that for an average plant the total cost of control per tonne of lead produced increased from £2 per tonne at 95% collection efficiency, to £4 per tonne at 99%, to £8 per tonne at 99.8% and £10 per tonne at 99.9%. Ultimately these costs have to be borne by the consumer in the form of higher prices, or shareholders and labour in the form of lower dividends and fewer jobs. Consequently there are important multiplier effects of environmental control which can clearly impact on important facets of the economy such as economic growth, unemployment and in some cases international competitiveness.

Benefits

A common factor that is encountered in the process of evaluation, is the frequent interdependence of benefits. This makes isolation of the effects of a particular strategy rather difficult. The interdependence can apply both to the uses of resources, and to the effects on those uses of different polluting agents. It is often impossible to distinguish the effect of one pollutant from another. This has been the case in measuring the impact of pollution on human health where the work of Lave and Seskin [13] is pre-eminent. They analyse the relationships between deaths from certain diseases, and the incidence of sulphur oxides and particulates, using a multiple regression on cross-sectional data. Whilst the results of studies such as these do have a certain credibility, correlations between the pollutants and such factors as climate and socio-economic status, may lead to multicollinearity. It is then possible, though perhaps not likely, that pollution may be only the 'proxy' for the real cause of the illness. An additional problem is posed by the need to value life itself. Ridker [14] treats this emotive subject by summing anticipated loss of earnings, the cost of premature burial, and medical costs. This is, of course, hopelessly inadequate, as must be any valuation which omits the psychic damage of ill health or death, but it may at least indicate the lower bound for the true value.

Ultimately an overall measure of benefit is what is required. One way of approaching this, is to conduct a survey which hopes to establish individuals' 'willingness to pay'. It must be said immediately that devising a survey which will generate useful results, is an extremely skilled task in any field. There is also the need to be sure of the honesty of the answers, particularly where the respondents believe these may affect public policy towards themselves. An additional problem which is specific to environmental surveys is the accuracy of respondents' perception of pollution's effects.

A second measure employed to estimate

Vol. 7
1979

benefits involved regressing wage levels across cities with different levels of air quality. The underlying assumption is that wage rate variations compensate for relative civic disamenities, which might be reasonable if the labour market was perfectly competitive. Since this is not the case, and for many industries wages are negotiated nationally, this approach is extremely unsatisfactory.

The National Academy of Sciences in the US [15] regressed real wage rates against various socio–economic variables together with the level of three air polluting agents, but the coefficients of the latter were, unsurprisingly, found to be insignificant because of multicollinearity. A further contribution [16] did manage to obtain significance by using fewer variables, but had an equally dubious theoretical basis.

One method of valuation which is more promising is the use of land or property values. Numerous studies have been undertaken which regress land or property values against socio-economic variables and indices of pollution, in order to derive a coefficient reflecting the effect at the margin on property values, of a particular increment in environmental quality. Quite apart from the problem of multicollinearity, there may be serious objections in principle to the use of this technique, which are most lucidly expressed in a paper by Polinsky and Rubinfield [17].

There are two stages involved in the technique. Firstly, a multiple regression equation is estimated in which property values are the dependent variable. Secondly, the values which would obtain after a particular change in environmental quality, *ceteribus paribus*, are calculated by the simple application of the relevant coefficient. The attendant implicit assumptions, are that the regression equation is capable of accurately predicting property value changes, and that these changes are a valid reflection of the 'willingness to pay' of society as a whole, for the improvement. Both of these can reasonably be called into question.

On the point of the prediction of future property values, the most obvious objection is that other factors will not remain constant, given a change in environmental quality. In particular, wages may be affected if there is an inflow of residents to the area, as might be transport costs. An even more serious obstacle may be provided by the second implicit assumption, that property value changes are a fair measure of overall benefit. This will only be true if all benefits accrue to property owners. This will clearly not always be the case.

Whilst a general equilibrium model would be necessary to evaluate benefits very accurately, Polinsky and Rubinfield suggest the interesting compromise of measuring only the immediate change in land values, before individual economic agents have had time to adjust to alterations in other factors. Thus it is clear that their analysis is strictly speaking only valid for marginal changes in environmental characteristics.

The first step of the analysis is to estimate the relationship between land prices on the one hand, and net wages, level of environment amenity and a level of utility which is constant for all members of a particular class, on the other hand. This is done separately for each class of individuals, by using observations from each area where any members of that class reside. This permits the identification of the utility function for each class, given that the form of the function has been correctly specified. Utility is defined as a function of land prices, net wages, and amenity. The next step is to identify the level of wages necessary to maintain the same level of utility, given an increase in amenity.

Since utility for all members of a class is the same, this cross-sectional approach avoids the problem of a potential income effect. The level of wage necessary to maintain equilibrium after an environmental improvement will be lower than the previous level. The final stage of the benefit calculation involves summing the reduction in wages over all locations, and for all relevant classes of consumer, to obtain an estimate of society's willingness to pay for the benefit.

Overall assessment of the use of cost–benefit analysis

It has been argued that the inclusion of the costs and benefits of environmental quality are essential to social decision-making, because almost all productive activities influence the environment in some way. Furthermore the residuals arising from these activities can only be transformed, not destroyed. In the context of the environment, the predominant role of cost–benefit analysis is to make assessments of

the value to society as a whole of alternative strategies towards the transformation of waste products, in terms of the benefits from environmental improvement and the costs from the usage of scarce resources.

The measuring rod of money is generally adequate so far as the valuation of the resource costs of achieving a particular environmental improvement is concerned, but is often inadequate as a yardstick for the magnitude of the benefit to society arising from the improvement. This is the explanation for the observable fact that so many of the larger scale cost–benefit studies undertaken in the field of the environment have concerned themselves with determining the most efficient or least cost means of attaining a particular level of environmental quality, rather than comparing alternative quality–cost configurations.

It would be wrong to infer from the undoubted existence of conceptual and practical difficulties in the evaluation of environmental benefits that no useful monetary estimates can ever be made of the benefit to society of a particular reduction in pollution. In some cases reasonable 'proxies' are available for benefit measurement, and in others it may be possible to assess the benefits by dividing them into components. Even where particular components of benefit are being estimated, valuations can only be made directly for material damage reduction and production benefits. In all other cases, values must be inferred from the behaviour of individuals.

So far as choices involving matters of aesthetic improvement or recreational provisions are concerned, there seems to be no reason for supposing that the preferences revealed by individuals are anything other than a true measure of the benefit derived. If, for example, 100 individuals would value the activity of fishing at a particular location at £10 each per annum, then it is unobjectionable to assert that the benefit from a pollution control device which prevents a pollution interference with the fishing is at least £1,000 per annum. Unfortunately the same assertion cannot be made for other categories of environmental benefit, in particular improvements in human health.

The basic problem arises from a lack of *understanding and knowledge* of the health effects of particular pollutants in differing concentrations. The effects on health of certain substances in the air and in water has been estimated by medical researchers, but by laymen most aspects of this subject are very poorly understood. A similar argument would apply to long-term damage of all types arising from the ecological implications of certain types of pollutant. For this reason, it is suggested here that the inference of values for human health from the observed behaviour of individuals is wholly inappropriate. The reasoning is not paternalistic in character. If members of society had sufficient knowledge and understanding of the complex technological interactions involved, then their revealed preferences would be an appropriate measure of benefit. *It is only because of this undeniable information deficiency that these preferences are not meaningful.* The implication of this must be that existing benefit evaluation relying on estimates of material damage or inferred behaviour must be replaced by a different type of measure.

Vol. 7
1979

AN ALTERNATIVE APPROACH

In response to this situation, a method has been developed by Battelle Columbus Laboratories [18] which uses an environmental index for the comparison of alternative pollution control strategies. This work was specifically directed at evaluating the environmental impacts of alternative methods of control, involving discharges to different media, since it is particularly difficult to express these comparisons in monetary terms. The aim of their method was to compare a number of possible environmental management alternatives for a particular productive process. An 'Environmental Degradation Index' (EDI) can be calculated for each alternative, and the 'Strategic Effectiveness Index' (SEI) of a particular control method was defined as the difference between its 'EDI' and the 'EDI' of the uncontrolled case (which would normally be higher). Thus:

$$SEI_i = EDI_u - EDI_i$$

where i is the relevant control strategy and u is the uncontrolled case.

The major innovation in this work was associated with the calculation of the appropriate 'EDI's', which necessarily involved a subjective relative valuation of emissions to different media. The approach to this was to develop

two levels of weights, the first for different pollutants within the same medium, and the second for comparing damage to alternative media. These weights would then be multiplied together to attain an overall weight for a particular pollutant. This weight was then subjected to an arbitrarily determined 'modifier' designed to take account of the characteristics of each pollutant in terms of its persistence, its geographical range of impact, and its transferability between media. The volumes of all emissions from a particular strategy were then applied to an arbitrarily determined 'S' shaped damage function based upon a hypothesised maximum level of emission of each pollutant. A damage index between 0 and 1 was obtained for each pollutant according to the volume emitted, and this was then multiplied by the weight appropriate to the particular pollutant. Finally the figure for all emissions were summed to arrive at the EDI for each strategy.

The development of the two levels of weights can be achieved by using a modified 'Delphi Technique'. Each member of a panel of experts was asked to rank a list of pollutants, and then to rate the second as a fraction in importance of the first, the third of the second, and so on. This process would be repeated separately for each industry to be considered, and comparisons made in terms of the maximum emission levels referred to above. An attempt to account for location was made by using regression data on the effect of certain key geographical factors on the environmental damage caused by pollutants. Consensus is then approached through further re-estimating stages.

The information derived from this technique falls short of that obtaining from a full cost–benefit analysis, insofar as every alternative will have two non-commensurable parameters attached to it, an Index of Strategic Effectiveness and a cost. Furthermore the use of the 'modifier' for the weightings means that the Index itself cannot be considered a truly cardinal measure. For this reason the method will not generally indicate an optimal solution, but where there are a number of alternatives under consideration, it will often point to some good solutions and some less good ones. This approach seems a most promising one and can be improved by some modifications which the authors have adopted in some recent research [19]. In particular the use of an arbitrarily determined 'modifier' to account for dispersion range, transferability, and persistence seems to run contrary to the subjective approach of the work. It might prove easier for experts to include these factors when making their valuations, than to deliberately exclude them in order for them to be accounted for separately. Allowing the experts to make the valuations inclusive of these factors also has the important advantage of retaining the cardinality of the index. Cardinality allows a comparison of alternatives on the basis of units of environmental benefit per £ of resource cost.

Secondly whilst the assumption of an S-shaped damage curve for pollutants is as plausible as any other assumption, it would be possible to extend the use of the 'Delphi' technique to include questions designed to find out the experts' own perception of the shape of the damage function for each pollutant.

Thirdly a more accurate picture of pollution problems can be obtained by modifying the 'Delphi' technique to allow the inclusion of brief statements of reasons for valuations by participants. These could then be circulated (anonymously) along with median results at the second stage. This would allow some limited debate without having the unfortunate disadvantages of personality dominance associated with face-to-face encounters.

Fourthly the use of a two stage weighting system seems to be both unnecessary, and fraught with difficulties. In particular, experts must have found it very hard to make comparisons of air, water and solid waste pollution in general terms. A straight relative weighting of relevant pollutants, whether to air, water or land, seems to be both more straigh'forward and more meaningful.

Although the relative effects of different pollutants are not couched in terms of money, it is important that participants have the concept of relative *economic* damage in their mind when making valuations, in order for the results to be meaningful. This should be emphasised in communications with the participants.

Finally a central problem which will always recur in an approach of this sort involves the question of location. The best way to proceed would seem to be to specify a hypothetical location by means of air and water quality parameters. The extent to which any results are

capable of generalisation will depend upon the extent to which a 'typical' urban location exists.

In our own research consensus on the rankings and weights applied to the individual pollutants increased considerably from the first round to the second round and this enabled an application of these essentially cardinal measures to alternative forms of environmental control. Rational choice between alternatives is thus facilitated although no measure of the absolute benefits or costs comparable to non environmental goods is possible—as it might be in a perfect cost–benefit study.

CONCLUSIONS

It has been argued that the inclusion of valuations of environmental quality are essential in government decision making because almost all productive activities influence the environment and residuals can only be transformed and not destroyed. Market and voluntary solutions are intuitively appealing but beset with practical problems. Here the role of economics in environmental decision making is philosophical and has important implications for justice and the rule of law.

Potentially cost–benefit analysis should enable adequate public decisions to be made although there are severe conceptual and practical limitations to its application. Consequently the use of indexes of environmental harm may be more appropriate in restricted choice situations. These types of solution are clearly preferable to arbitrarily determined ones although are not as powerful as those provided by ideal cost–benefit analysis.

REFERENCES

1. Joint Working Party on Lead Pollution around Gravelly Hill (1978) Lead pollution in Birmingham. *Dept of Envir. Publ. Pap.* **14**, HMSO.
2. Atomic Energy Research Establishment (1978) *Investigation into Lead from Motor Vehicles.* HMSO.
3. Ayres R and Kneese AV (1969) Production, consumption and externalities. *Am. Econ. Rev.* **59**, 282–297.
4. Coase R (1960) The problem of social cost. *Jl Law & Econ.* **III**, 1–44.
5. Dick DT (1976) The voluntary approach to externality problems: a survey of the critics. *Jl Envir. Econ & Mgmt*, **2**(3), 185–195.
6. Pearce DW (1976) The limits of cost–benefit analysis as a guide to environmental policy. *Kyklos*, **29**, 97–112.
7. Lowe JF & Atkins MH (1978) Resource conservation and environmental control. *Resour. Policy*, **4**(1), 45–52.
8. Gray PMJ (1976) Design requirements for emission control in non ferrous extraction plant. In *Engineering Aspects of Pollution Control in the Metals Industry.* Metals Society.
9. Lowe J (1978) *The economics of clean air.* An. Conf. of Nat. Soc. for Clean Air. **45**.
10. Atkins MH & Lowe JF (1979) *Cases in Pollution Control in Textile Dyeing and Finishing.* Pergamon Press, Oxford, UK.
11. Atkins MH & Lowe JF (1977) *Pollution Control Costs in Industry: an Economic Study.* Pergamon Press, Oxford, UK.
12. Atkins MH & Lowe JF (1979) *The Economics of Pollution Control in the Non Ferrous Metals Industry.* Pergamon Press, Oxford, UK.
13. Lave L & Seskin EP (1970) Air pollution and human health. *Science.* **161**, 723–753.
14. Ridker RG (1967) *Economic Costs of Air Pollution.* Praeger Press.
15. National Academy of Sciences (1974) The cost and benefits of automobile emission control. *Air Qual. & Emission Control*, **4**,
16. Nordhams WD & Tobin J (1973) Is growth obsolete? In *The Measurement of Economic and Social Performance* (Ed. Moss), National Bureau of Economic Research.
17. Polinsky AM & Rubinfield DL (1975) Property values and the benefits of environmental improvements: theory and measurement. *Harvard Discuss. Pap.* **404**.
18. Battelle Columbus Laboratories (1974) *Final report on the development of cross media evaluation methodology.* EPA Publ. No. PB-232-414.
19. Lowe JF, Lewis D & Atkins M (1979) *The environmental consequences of pollution control.* Report to the Department of the Environment.

ADDRESS FOR CORRESPONDENCE: *J Lowe, Esq, School of Management, University of Bath, Claverton Down, Bath, BA2 7AY, UK.*

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 5, pp. 431 to 439
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1001-0431$02.00/0

Environmental Impact Statements

M BARRETT

University of Aberdeen, Scotland

(*Received May* 1979)

Environmental Impact Statements (EISs) present the results of a structured appraisal of the likely effects of a proposed development on the environment. Systematic methods of identification and assessment of potential effects of major projects have been developed in response to public concern about environmental quality. EISs were first introduced in the United States (US) in 1970 but they have since been adopted in a number of countries. In the United Kingdom (UK) EISs have been prepared for certain major projects, in particular for developments related to North Sea oil, but the procedure has not been formally introduced into the planning system. In this article the history of EIS in the US, UK and elsewhere is traced. Contents of EISs and methods of appraisal which may be included are described. Many attitudes towards EIS are based on the early years of EIS in the US where it was seen to be a slow, costly procedure frequently involving lengthy court battles. A review of EIS experience in the US shows that these problems have been exaggerated. Two major developers in the UK, BP and British Gas Corporation, consider EIS to be a valuable component of their project planning activities, ensuring that environmental aspects of developments are considered at an early stage. Finally some arguments for and against the formal introduction of EIS in the UK are examined. The article concludes that EIS will play an increasingly important role in the UK planning system.

INTRODUCTION

IN RECENT years there has been increasing public concern about the quality of the environment and consequently about the effects of major developments on the environment. The 'environmental movement' first gained support in the United States (US) in the 1960s but since then its influence has spread to most Western countries. The impact of a proposed development on the environment can now be a critical consideration in determining whether or not it should proceed. Methods of predicting, appraising and presenting the environmental impacts of proposals have been developed to meet the needs of decision-makers for information on the environmental implications of their decisions.

One response to the growth of interest in environmental conservation in the US, was the passing of the National Environmental Policy Act (NEPA) in 1970 [23]. NEPA introduced the concept of environmental impact statements (EISs). An EIS is a document which sets out the likely impacts of a proposed project, plan or policy on the environment. It presents the results of systematic assessment of the potential effects of a proposal. Impact assessment is a prerequisite to the production of an EIS. Throughout this article the term EIS refers to the process of assessment of likely impacts and to the statement of outcome of assessment. EIS is a structured approach to environmental appraisal of projects. It makes use of procedures and methods of assessment have been developed to facilitate systematic appraisal.

Since the enactment of NEPA, interest has been expressed in EIS in many countries and a number have introduced procedures which require the preparation of EISs for certain types of projects. In the United Kingdom (UK) the question of whether EIS should be formally incorporated into town and country planning legislation is unresolved. It has been the subject of government sponsored research and is cur-

rently being debated widely in planning and academic circles. This article traces the history of EIS in the US, the UK and elsewhere. The contents of EISs and methods of identification and assessment of impacts which they may include are described. The current controversy about EIS in the UK is examined in the light of experience of EIS in the US and the attitudes of selected developers towards EIS.

HISTORY OF EIS

The enactment of NEPA on 1st January 1970 marked the first legal recognition of EIS. NEPA declared a US national policy of encouraging harmony between man and the environment. The means of achieving this policy objective were contained in the 'action-forcing' provisions of the second part of the Act.

The most important of these required Federal agencies to produce EISs for "major Federal actions significantly affecting the quality of the human environment". The EIS requirement is limited to public projects and policies and to those private developments for which Federal permits, authorisation or funding is necessary. NEPA introduced EIS as an administrative procedure to ensure that Federal agencies complied with national environmental policies. It was not intended as a land-use planning tool as it has later come to be considered in the UK and elsewhere.

The ambiguity of NEPA, which used but did not define such terms as 'major' and 'significantly', resulted in many court actions in the years immediately following its implementation. Prolonged legal battles about EISs damaged the reputation of the procedure especially amongst observers outside the US. However, litigation owed more to the drafting of NEPA, the inexperience of many Federal agencies in environmental matters and the legal system in the US than to the intrinsic characteristics of EIS. The possibility of court action led Federal agencies to produce voluminous EISs which covered every conceivable impact, a notorious example being the nine-volume EIS on the Trans-Alaska Pipeline [27]. The US Council on Environmental Quality (CEQ), which is responsible for overseeing the implementation of NEPA, has recently issued regulations governing the preparation of EISs. Among other things the need for brevity is stressed, requiring that an EIS prepared under NEPA should not normally exceed 150 pages [26].

The discovery of North Sea oil and gas was an important factor in the growth of interest in EIS in the UK. Exploitation of the oil and gas required large-scale developments, such as platform construction yards, which placed quite new demands on the planning system. The Scottish Development Department (SDD) responded to the special problems of oil-related development by advising planning authorities that such developments ought to be subject to "rigorous appraisal" [19]. Planning authority in-house expertise was in many cases unable to provide such an appraisal. To ensure effective assessment, SDD chose to call-in major oil-related development applications for decision by the Secretary of State and to employ consultants to produce EISs. SDD was consequently in the vanguard of research into EIS in the UK.

In 1974, the Secretaries of State for the Environment, Scotland and Wales commissioned an investigation of the desirability of introducing a system of EIS in Great Britain. The study concluded that a limited number of development proposals (possibly 25–50 a year) which had large scale and complex impacts on the environment ought to be subject to EISs [6]. A similar conclusion was reached in a Review of the Development Control System published in 1975 [9]. Catlow and Thirlwall recommended that EISs could be introduced experimentally without legislation but that a number of amendments to the Planning Acts would be desirable if EIS were to become a ermanent part of the UK planning system. These included giving the appropriate Secretaries of State the power to prescribe types of development which should be subject to an EIS and providing that the costs of an EIS should be shared between the planning authority and developer. The Department of the Environment published the report (1979) without commenting on its conclusions and there has still been no official response from central government.

The Project Appraisal for Development Control (PADC) research team of the Geography Department, Aberdeen University, was

Vol. 7
1979

sponsored by central government to develop a method which enabled planning authorities to make a balanced assessment of the potential impacts of industrial developments. The method was published in *The Assessment of Major Industrial Applications—A Manual* [17]. The primary purpose was to aid planning authorities in assessing applications for planning permission for private developments, although the approach can be applied to public sector projects. At present the Manual is being revised taking account of the experience of those who have used it. A revised and extended Manual will be published toward the end of 1979.

Despite commissioning these studies, the UK central government has been criticised for its lukewarm attitude to EIS [4]. In the present climate of constraints on local government expenditure there has been no move to introduce EIS as a mandatory addition to planning authority functions. The government's industrial strategy accords high priority to the encouragement of industrial development and any new procedure which may hamper industrial growth is treated with extreme caution.

The introduction of EIS may be viewed by the government as being incompatible with the industrial strategy. A few planning authorities have prepared EIS for exceptional developments but more frequently environmental consultants have been employed to undertake assessments.

While moves towards the introduction of EIS by the UK government have been cautious, other countries have adopted EIS procedures. In 1973, Canada introduced EIS for major projects planned or funded by Federal agencies. The Canadian procedure requires that an independent Environmental Assessment Panel is constituted for each development to define the scope of appraisal, review the assessment and make recommendations to the appropriate Minister. Australia implemented legislation requiring EISs for Federal developments in 1975. In Europe, France and Eire have enacted EIS legislation while West Germany has instituted dministrative procedures covering the preparation of EISs. Within a decade EIS has become part of the project assessment and authorisation machinery in many countries.

THE NATURE AND CONTENT OF EISs

An EIS presents the results of an assessment of the likely impacts of a proposed development to the public and to those responsible for deciding whether the development will proceed. EIS has the dual functions of 'opening up' decisions to public scrutiny and of improving the factual basis for decisions. The purpose of EISs is, therefore, to communicate the implications of possible actions to non-experts.

There is no standard form for an EIS. NEPA requires that the following aspects should be discussed:

(i) the environmental impacts of the proposed action;

(ii) any adverse environmental impacts which cannot be avoided;

(iii) alternatives to the proposed action;

(iv) the relationship between local, short term use of man's resources and the maintenance and enhancement of long term productivity;

(v) any irreversible and irretrievable commitment of resources.

Many EISs in the United States follow these headings although extra sections are often added which cover description of the project, description of the site and surrounding area and possible actions to mitigate adverse effects. A discussion of mitigating actions has proved to be a valuable component, indicating to decision-makers how projects can be made more acceptable. Canadian EISs are not constrained by legal requirements and show a wider range of contents.

The PADC Manual suggests a pattern for EIS which is partly based on American EISs but which takes account of the needs of a planning authority determining an application for planning permission; six components are suggested:

(i) developers proposals;

(ii) description of the local area;

(iii) potential impacts;

(iv) mitigating actions;

(v) effects of existing trends;

(vi) consultations and objections.

The first section includes more than a simple description of the development, providing information about the developer, his reasons for selecting the site, other sites considered, infrastructure and emergency service requirements. The description of the local area provides the baseline information from which changes in the environment are predicted and may subsequently be measured. The remaining sections present the results of analysis of likely impacts of the development and predictions for the future of the area should the development not take place. This framework has been drawn up for use within the present UK planning system and provides for inclusion of all material considerations for determining planning applications.

In the early 1970s most EISs prepared under NEPA concentrated on impacts on the natural environment. CEQ guidelines published in 1973 emphasised the natural environment but recommended inclusion of social and economic impacts which could be measured quantitatively [24]. Consequently, EISs usually contained discussions of demographic changes, effects on employment opportunities and the financial implications of the project. Less easily measured impacts such as effects on community cohesion, social relationships and political and economic structures were rarely considered.

One difficulty of introducing social impacts into EISs is that techniques for predicting social change are less well established than those for environmental impact prediction. In the field of social change it is necessary to rely heavily on the experience and knowledge of experts and anticipated impacts may appear to be less 'scientific' than ecological predictions. However, social and economic impacts will almost always be significant and may be of overriding importance in determining whether a development should proceed and an EIS benefits from covering them as comprehensively as possible.

Methods for impact appraisal serve one or more of the following functions:

(i) to identify impacts, including indirect impacts;

(ii) to provide a framework for structured appraisal of impacts;

(iii) to assess the importance of impacts;

(iv) to present information on the impacts to decision-makers and to the public.

A variety of methods have developed which range from simple checklists to complex modelling approaches. A common method is based on matrices which relate actions involved in developing the project to components of the environment. Where an action is expected to have an impact on a feature of the environment the cell in the matrix representing the interaction is marked. On the basis of a simple interaction matrix of this type, a more complex matrix method was developed by the US Geological Survey [14]. In this matrix scores for the magnitude and importance of each impact are inserted in the relevant cell.

A limitation of Leopold-type matrices is that they cannot be used to identify second and higher order interactions [1]. A matrix designed by Environment Canada traces indirect impacts using a matrix multiplication technique [10]. However, extremely complex calculations are required and it is doubtful if these are justified by the results [18]. The Leopold matrix is open to a criticism which can be levelled at all methods which assign ordinal values to impacts. The scores for importance and magnitude are determined subjectively by experts and are not comparable between impacts. Nevertheless, there is a temptation to add scores to produce an aggregate impact value which leads to a false 'result'. The use of numerical scores disguises the qualitative nature of impacts and may remove much of the relevant information from the public and decision-makers.

Matrices can be laborious to use because all possible interactions have to be checked. But, despite their drawbacks, matrices are the most widely used methods of impact identification. They act as useful aid-memoires for those assessing impacts and are a good visual means of communicating information about likely impacts.

Network analysis and systems diagrams have been employed to identify second and

Vol. 7
1979

higher order impacts. A network developed by Sorensen traces the final effect of a development action through a series of 'condition changes' in environmental parameters [20]. This method identifies indirect impacts but does not attribute any measure of their relative importance. Networks and systems diagrams can become complex when applied to projects with a large number of impacts and access to a computer with an appropriate data base may be necessary. The US Forest Service found networks useful in assessing forest developments [22].

In the method developed by PADC, a series of linked activities begin with informal discussions with the developer and end at the final decision of the planning authority. Acquisition of data on the project, the site and surrounding area, consultations and public involvement, completion of an impact matrix and appraisal of likely impacts precede the production of an EIS. The information gained from all these activities contributes to the preparation of the EIS which in turn forms the basis of the Major Issues Report. The report concentrates on the important impacts and provides supporting argument for the decision.

EXPERIENCE OF EIS IN THE US

Since the passing of NEPA much experience of the problems and value of EIS has been gained in the US which is relevant to the debate about the introduction of an EIS procedure in the UK. The lessons cannot be translated directly, however, because many of the characteristics of EIS in the US arise directly from the administrative and legal systems in that country. The NEPA procedure also varies considerably from any system that would be likely to be introduced in the UK. NEPA provides that Federal agencies should prepare an EIS before proceeding with any major action. The body proposing the development is, therefore, also responsible for the EIS. Comparable arrangements in the UK would require government departments and statutory undertakers such as the Central Electricity Generating Board to produce and publicise EISs for major developments. The primary purpose of EIS in the UK, however, is generally seen to be as an aid to planning authorities exercising their development control functions. Bearing in mind these differences in the procedure and function of EISs in the two countries, a review of the experience in the US indicates the benefits and disadvantages EISs may bring.

The cost to agencies of compliance with NEPA is central to most discussions of EIS. CEQ, in reviewing the first 6 years experience of NEPA, stated that it was not possible and may never be possible to determine the exact costs of EIS preparation [25]. This is principally because the intention behind NEPA was to integrate environmental appraisal with other project planning activities and where this integration is achieved it is impossible to isolate EIS costs. However, Federal agencies were asked by CEQ to estimate expenditure on the preparation and review of EISs. These estimates showed that EIS costs were not high compared with the operating budgets of agencies or with the cost of major projects. For example, the Corps of Engineers spent about $21.0 million on EISs in 1974 which was approximately 1.2% of its annual budget [25]. A more recent figure calculated by the Controller General's Office on the basis of 29 case studies was that the average cost of preparing an EIS amounted to 0.01% of total project costs [21]. The costs of EISs have not, therefore, proved to be prohibitively high compared with other project expenses.

Another view of the cost of EIS is that the true price lies in the delay in commencing developments. This attitude may have particular force in the UK where industrial regeneration is a high priority. In 1976 CEQ published average times taken to prepare draft EISs. The averages varied considerably from as little as 3 months for the Department of Housing and Urban Development to 20 months for the Bureau of Land Management. More detailed information from the US Atomic Energy Commission shows that the average time taken to complete a draft EIS was 8½ months or 420 man days. The period could be reduced by increasing the staff; the minimum time taken was 5 months or 640 man days [12]. In the early years following enactment of NEPA, EIS gained a reputation for creating very long delays. This was mainly due to the large backlog of projects for which EISs were required because the provisions of the Act applied retrospectively. As the backlog was cleared and

experience in the preparation of EISs was acquired, the delays were reduced.

A source of delay has been litigation arising out of NEPA. During the 5 years 1970–75 approximately 5% of EISs were challenged in the courts. In the UK, the courts do not become involved in the merits of planning matters; an objector to a planning decision has recourse to the courts only on a point of law and cannot challenge the decision itself. Criticisms of NEPA based on the amount of litigation it generated are not directly applicable to the UK. Public concern about environmental issues in the US may have found expression in lengthy court actions attempting to block controversial developments even if NEPA had not been enacted.

There is no simple way of judging the effectiveness of NEPA in reducing the harmful environmental effects of Federal actions. Few projects have been abandoned because an EIS demonstrated that they would be environmentally deleterious [18]. Where environmental appraisal is integrated with other project planning activities, the development may be modified to take account of environmental considerations before an EIS is prepared. It has been suggested that in many cases environmental factors uncovered in the course of preparing EISs have caused the projects to be altered [15]. As Federal agencies are responsible for originating developments, undertaking environmental appraisals and deciding whether developments should proceed, there is no external agency to ensure that decision-makers make full use of EISs. The influence of EISs on the internal decision-making of an agency is difficult to measure. The Environmental Protection Agency is currently investigating the effectiveness of its own EISs and holds the view that the only meaningful way to evaluate EISs is by in-depth case studies [7].

An indirect but possibly very significant effect of NEPA is the raising of environmental awareness of agency staff and the public. The knowledge that an EIS will be required before a project may proceed means that environmental factors, even when they are not explicitly considered, are always in the background of project planning. Ecologists and other environmental specialists employed to complete EISs may affect attitudes within agencies; the Corps of Engineers, for example, employs more than 197 environmentalists assigned entirely to NEPA activities [16].

The purpose of NEPA is to ensure that Federal agencies carry out their functions with minimum adverse effects on the environment; the EIS procedure is a means to this end. The value of EIS lies both in its direct and indirect influences on decision-makers. EIS has been criticised as having become a matter of expensive routine often producing lengthy and dull documents [11]. Defenders of NEPA claim that while such criticisms may be justified, they reflect the misuse of the procedure and not the shortcomings of the Act itself [5]. Problems of cost and delay associated with EISs in the US which have created some resistance to the introduction of EIS in the UK, appear to have been exaggerated. These problems are consequences of the political and judicial systems in the US and may be avoided in the UK.

INDUSTRIALISTS' ATTITUDES TO EIS

Although the main purpose of EIS is to prevent developments taking place which would unacceptably harm the environment, some industrialists have found that EIS can serve their interests. In a recent interview Jack Birks, the Managing Director of BP, stated that "One thing we have learned from our international experience is the value of the EIS, anticipating well ahead of the activity what the environmental consequences might be". He went on to say that an early Shetland EIS was invaluable in enabling BP to proceed with the Sullom Voe oil terminal [2].

BP has an internal formal system of EIS for all major new developments. An EIS is prepared by the associate or subsidiary company proposing the development and submitted to the Group's Central Environmental Control Centre. Endorsement by the Environmental Control Centre is required for any planned capital investment of more than £5m. It is claimed that environmental considerations are taken into account at every stage of design of a major development and in advance of a planning application being made. BP benefits from speedier processing of planning applications, from avoidance of unnecessary conflicts on conservation grounds, and from the reduced

risk of delay or disruption arising from lengthy and costly public hearings [3].

The British Gas Industry has also found EIS to be a valuable tool in project planning and has developed a number of methods of appraisal, particularly concentrating on visual impact analysis. Frank Deans, Chief Environmental Planning Officer of the British Gas Corporation accepts that the preparation of an EIS is a time consuming and expensive operation but believes that it produces significant advantages. "Firstly, it is unquestionable that in the end, a much better designed and sited installation will result and this will have a long term financial advantage. Secondly, the increasing awareness of the general public of the need for care and protection of the environment and the growing practice of planning authorities of involving the public in the decision-making process could and does lead to serious delays in obtaining the necessary consents. If an environmental impact analysis has been carried out, it will demonstrate to all that the decision to develop at a certain place and in a certain manner has not been lightly taken and this should reduce the overall time for the project to be completed, with a consequent saving in costs" [8].

Both BP and British Gas recognise the value of EIS as an inseparable part of project planning which ensures that environmental factors are given full weight. By integrating the environmental considerations at the outset, objections and delays at later stages may be avoided. The preparation of an EIS by a planning authority on receipt of a planning application serves a different purpose of informing decision-makers and the public. Nevertheless, the same advantages of undertaking a comprehensive appraisal as early as possible to preempt later delays apply. An EIS may reveal an adverse impact that can be avoided by modifying the proposal but which were it not to come to light until a later stage, perhaps at a public inquiry, would cause significant delays.

THE FUTURE OF EIS IN THE UK

Informal use of EIS has been made by developers, planning authorities and by UK central government but disagreement remains about the introduction of a formal EIS system. One argument which is made against the introduction of such a system is that it is not needed. It is suggested that countries without comprehensive planning controls require special safeguards against possible detrimental effects of major developments but that in the UK such safeguards already exist. UK planning and pollution controls combine to provide a very thorough system of authorisations for major developments. Nevertheless, there is a body of opinion which holds that the introduction of EIS would strengthen and improve control [6, 9]. Support for the introduction of a system of EIS has broadened in the last year or two, largely because of doubt about the ability of established planning procedures to deal with exceptional developments such as the proposal to expand reprocessing of nuclear fuels at Windscale. The value of EIS to provide thorough appraisal of all aspects of major proposals beyond the narrow planning issues has been recognised by professional bodies such as the Institute of Civil Engineers and the Town and Country Planning Association.

The flexibility of the UK planning and pollution control system which is its strength provides an argument for the introduction of EIS. Appraisal of a planning application may be as rigorous or as cursory as the planning authority considers necessary as long as all material considerations are taken into account in reaching the decision. An EIS system would ensure a degree of consistency of treatment of the few most significant projects throughout the country.

Discussions about the role of EIS in the UK have been given added impetus by the preparation of a preliminary draft directive by the European Commission proposing a mandatory system of EIS for all member states. The preliminary draft directive is based on the recommendations of consultants appointed by the Commission [13]. It proposes that authorisations or approvals for development projects which are likely to have a significant impact on the environment should be preceded by an EIS. A list of types of projects which will be subject to EIS procedure is included. The draft is now being discussed by member state governments and may be altered considerably before being adopted. It seems likely that government departments who sponsor major developments

such as energy undertakings, will be wary of additional control which may complicate and delay authorising procedures. Member states with well established land use planning systems may fear that a compulsory Community EIS process will not integrate well with existing arrangements.

Despite the cautious attitude of the government towards EIS it appears likely that it will become a feature of the UK planning system in the next few years. A number of developers and professional organisations accept that major developments with far-reaching environmental effects require rigorous appraisal which can best be accomplished by an EIS. EIS may be introduced informally on a wider scale than at present without any changes to planning legislation. The initiative of the European Commission may, however, result in the formal adoption of EIS. EIS would strengthen land-use planning controls in the UK by providing a procedure for dealing with exceptional developments which could not have been foreseen in 1947 when the planning system was established.

CONCLUSIONS

With increasing public interest in environmental protection, the impact of a proposed development on the environment has become a significant factor in determining whether or not it will succeed. Projects which are expected to have a detrimental effect may be delayed by objections and in the end may have to be abandoned. The need for more information about the effects of major installations on the environment has led to the development of procedures of impact appraisal. EIS may be used by authorising or funding agencies to determine whether developments should proceed. In these circumstances the appraisal would be undertaken when the planning of the project is well advanced. EIS also has a role earlier in the project planning process when used by developers to ensure that environmental issues are considered at an early stage. The image of EIS as a slow, bureaucratic procedure is largely based on the years immediately following the implementation of NEPA in the US. As experience has been gained in the US and elsewhere, the value of EIS as a planning tool has been recognised. In response to demands for more effective assessment of major developments, EIS is likely to play an increasingly important role in the UK planning system.

ACKNOWLEDGEMENT

This article is based on work undertaken by the author on a project sponsored by the Scottish Development Department, Department of the Environment and the Welsh Office. The views expressed are the author's and do not necessarily reflect those of the sponsoring departments. I am grateful to Ms Pearl Allan for typing the manuscript.

REFERENCES

1. ANDREWS RNL (1973) *Approaches to impact assessment: comparison and critique.* Paper presented to the Short Course on Impact Assessment in Water Resources Planning, Ann Arbor, Michigan, 9 June 1972.
2. BIRKS J (1979) Interview. *J. Petrol. Technol.*, **31**(3), 287–290.
3. BP (1979) *Briefing paper on environmental protection* ENDS **20**, 19–23.
4. BRINSFORD J (1979) EIS filibuster. *Planning*, **304**, 9–10.
5. CALDWELL LK (1978) The environmental impact statement: a misused tool. In *Environmental Impact Analysis: Emerging Issues in Planning* (Eds JAIN RK & HUTCHINGS BL). University of Illinois Press, Urbana, Illinois, USA.
6. CATLOW J & THIRLWALL G (1977) *Environmental impact analysis.* Dept. of Envir. Res. Rep. **11**, Department of the Environment, London, UK.
7. COOK PL (1978) Practical problems in implementing NEPA in the Federal bureaucracy. In *Environmental Impact Analysis: Emerging Issues in Planning* (Eds JAIN RK & HUTCHINGS BL) pp. 27–36. University of Illinois Press, Urbana, Illinois, USA.
8. DEAN FE & GRAHAM G (1977) *Environmental impact analysis in the British Gas Industry.* Paper presented to the Sympsoium on the Gas Industry and Environment, Minsk, USSR, 20–27 June 1977.
9. DOBRY G (1975) *Review of the development control system*, HMSO, London, UK.
10. ENVIRONMENT CANADA (1974) *An Environmental Assessment of Nanaimo Port Alternatives.* Environment Canada, Ottawa, Canada.
11. GARNER JF (1979) Environmental impact statements in US and UK. *J. Plann. & Envir. Law.*
12. HOGLUND BM (1976) Environmental impact statements—their preparation and their problems. *Environmental Impact of Nuclear Power Plants* (Eds KARAM RA & MORGAN KZ). pp. 446–458. Pergamon, New York, USA.
13. LEE N & WOOD C (1976) *The introduction of environmental impact statements in the European communities.* Document **ENV/197/676.** Commission of European Communities, Brussels, Belgium.
14. LEOPOLD L, CLARKE FR, HANSHAW BB & BALSLEY JR (1971) *A procedure for evaluating environmental impact.* US Geological Survey Circular **645**, US Geological Survey, Washington DC, USA.
15. MEYERS S (1976) US experience with national environmental impact legislation. In *Environmental Impact Assessment* (Eds O'RIORDAN T & HEY RD) pp. 45–55. Saxon House, Farnborough, UK.

16. OTTO AC (1978) P and S integration—corps implementation. In *Environmental Impact Analysis: Emerging Issues in Planning* (Eds JAIN RK & HUTCHINGS BL) pp. 63–71. University of Illinois Press, Urbana, Illinois, USA.
17. PROJECT APPRAISAL FOR DEVELOPMENT CONTROL (1976) *Assessment of major industrial applications: a manual.* Dept. of Envir. Res. Rep. **13.** Department of the Environment, London, UK.
18. PROJECT APPRAISAL FOR DEVELOPMENT CONTROL (1978) *Environmental impact assessment in the USA: a critical review.* Dept. of Envir. Res. Rep. **26.** Department of the Environment, London, UK.
19. SCOTTISH DEVELOPMENT DEPARTMENT (1974) *Appraisal of the impact of oil-related development.* **DP/TAN/16** Scottish Development Department, Edinburgh, Scotland, UK.
20. SORENSEN JC (1971) *A framework for identification and control of resource degradation and conflict in multiple use of the coastal zone.* Master's Thesis, Department of Landscape Architecture, University of California, Berkeley, USA.
21. SUSSKIND L (1978) It's time to shift our attention from impact assessment to strategies for resolving environmental disputes. *EIA Rev.* **2,** 4–8.
22. THOR EL, ELSNER GH, TRAVIS MR, O'LOUGHLIN T (1978) Forest environmental impact analysis—a new approach. *J For.* **76,** 723–725.
23. UNITED STATES CONGRESS (1969) National Environmental Policy Act of 1969. *Public Law* **91-190,** 1–5.
24. UNITED STATES COUNCIL ON ENVIRONMENTAL QUALITY (1973) Preparation of environmental impact statements: guidelines. *Fed. Reg.* **38,** 7724–7729.
25. UNITED STATES COUNCIL ON ENVIRONMENTAL QUALITY (1976) *Environmental Impact Statements: An Analysis of Six Years' Experience by Seventy Federal Agencies.* Council on Environmental Quality, Washington DC, USA.
26. UNITED STATES COUNCIL ON ENVIRONMENTAL QUALITY (1978) National Environmental Policy Act: proposed regulations for implementing procedural provisions. *Fed. Reg.* **43,** 25230–25247.
27. UNITED STATES DEPARTMENT OF THE INTERIOR (1972) *Final Environmental Impact Statement on the Proposed Trans-Alaska Pipeline.* Department of the Interior, Washington DC, USA.

ADDRESS FOR CORRESPONDENCE: *Ms M Barrett, Project Appraisal for Development Control, Department of Geography, University of Aberdeen, Aberdeen AB9 2UF, Scotland, UK.*

1. 7
79

BLANK
PAGE

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 5, pp. 441 to 450
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1001-0441$02.00/0

The Use of Economic Appraisal in Transportation Analysis

CA NASH

University of Leeds, UK

(*Received May* 1979)

This paper surveys the use of economic appraisal techniques in the field of transportation. After surveying the history of such approaches, a critical assessment is offered of the most recent contributions to the debate, centering on the 1977 Leitch Report. The limited room for manoeuvre in making local assessments is highlighted: national policy considerations are seen as frequently dominating the local choice. While most of the debate on the 'inadequacies' of cost–benefit have focussed on 'intangible' items, the author argues for a much closer look at the 'hard' data on the value of time-savings and accident costs. Some of the Leitch committee's proposals are found wanting in this respect. In the wider context, a plea is made for the use of common criteria for assessing projects throughout the entire transport sector. Only in this way can misallocation of resources be minimised.

BACKGROUND

THE EARLIEST paper to which students of the use of economic analysis in the transport sector are usually referred was published by a French engineer, Dupuit, in 1844 [10]. He posed the problem of how to value the output of public works such as bridges, canals, roads or railways which generate a substantial increase in the volume of traffic passing between two points. To value the increase at the cost of transport by the previous method would clearly overstate the value of the output, since at that rate the traffic did not pass. To value it solely at the new cost ignored the fact that much of the traffic would be willing to pay more than that amount to pass. The solution used was to value the traffic at what the user was willing to pay to use the facility, as measured by the area under the demand curve for the facility in question, (Fig. 1). The foundations of the technique of economic appraisal of transport projects as in use today had been laid.

It took a long time to move from this beginning to the widespread use of the technique in practice. In the US, cost–benefit analysis came to be used to examining water resource projects in the 1930's, but it was developed along pragmatic lines without much thought being given to the theoretical background. In the UK, the pioneering studies were those of the M1 motorway [3] and the Victoria Line Underground railway [12]. From the M1 study approach was developed the methodology for the first use of the technique on a regular basis for evaluating trunk road proposals [8]. This approach was further refined and computerised in the form of the COBA programme in use today [16, ch. 4]. Following on the Victoria Line study, the technique began to be used in the evaluation of public transport proposals where these were believed to confer substantial benefits on persons other than users of the system (mainly continuing road users, who benefitted from a reduction in congestion brought about by diversion of some motorists to public transport). In the White Paper on *Transport in London* [5, Appendix J], a shortcut method of evaluation of such schemes was put forward, and this was the basis of the evaluation of a number of schemes in the years 1968–74, for which operators were applying for infrastructure grants under the 1968 Transport Act.

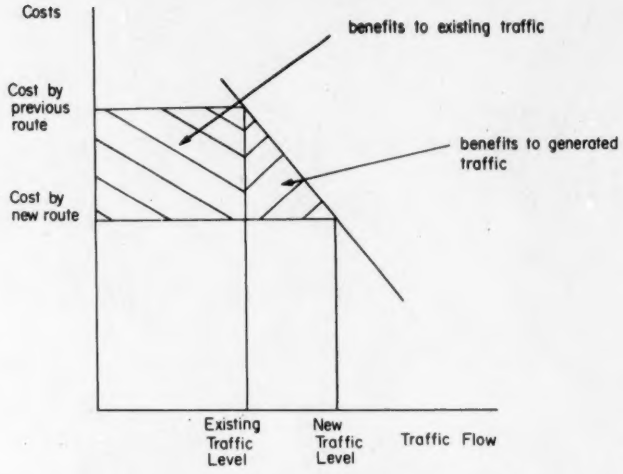


FIG. 1. *The basic framework of user benefit estimation.*

In the meantime, the major British conurbations were subjected to comprehensive land-use/transportation studies, designed to forecast and aid planning 20–30 years ahead. This approach was pioneered in the US and initially took the form of producing a set of forecasts for the design year, and designing a network to cater for this traffic, with little or no attention to economic appraisal [19]. Gradually, techniques were formulated for the economic appraisal of the alternative networks tested, until by the time of the Greater London Development Plan public hearing, these techniques were the subject of fairly sophisticated economic debate [14].

Up to this point, progress in the story has been essentially one way, the gradual spread and refinement of the use of economic appraisal in transportation analysis. In the last few years, however, there has been a growing disillusion in some quarters with the whole approach. In part, this arises from a suspicion that existing techniques can be, and are, manipulated to give whatever answer the body in question wants, and partly from a belief that with the best will in the world, the methodology cannot cope with the kind of problems that are involved in transport decision taking today. The turning point is often suggested to have come with the publication of the report into the siting of the Third London Airport in 1970. This study, costing more than a million pounds, was certainly the largest ever undertaken into a single transport project in this country. It aroused widespread disagreement as to the appropriateness of the methodology [20, 23], and the recommendation of the Commission of Inquiry was over-turned by the government. Whilst the latter in itself is no great criticism of the study (the study may have yielded the right information, but the government may have weighed it differently in taking the final decision, or the government may simply have taken the decision on political grounds not encompassed by economic appraisal), the former criticisms are a great deal more serious.

At the same time, objectors to new road proposals at public enquiries were beginning to examine official appraisals of the schemes more closely, and to conclude that all was not well there either. In the first place, the traffic forecasts which justified a large increase in capacity on many routes, and which were based on a simplistic approach to forecasting growth in car ownership, were questioned [1]. Secondly, the examination of alternatives (alternative routes and capacities for new roads, ways of relieving pressure for the road by improvement of public transport) was often thought to be inadequate. But thirdly, and perhaps most damaging of all, the techniques of appraisal were felt to place far too much weight on quantifiable, but not necessarily terribly important, time savings whilst giving little consideration to the quality of the environment in the area concerned.

It was this position which led the government to set up, in December 1976, an Advisory Committee on Trunk Road Assessment, whose

Vol. 7
1979

terms of reference were to review techniques of traffic forecasting, appraisal of trunk roads (taking into account both economic and environmental factors) and the problem of comparability of investment techniques between modes. This Committee (the Leitch Committee) reported in October, 1977 [18]. Whilst it considered that the appraisal techniques used by the Department were basically sound as far as they went, it made a number of important recommendations. Firstly, it considered the existing approach to be unbalanced and suggested a framework within which environmental and distributional factors should be considered alongside the existing economic factors. Secondly, it believed that existing techniques of traffic forecasting should be sensitive to alternative policies and that the range of uncertainty in the forecasts should be explored more fully. Thirdly, it believed that there was a need for further efforts to ensure compatibility between national and local roads programmes, and that strategic and policy studies should be undertaken to compare the cost–benefit rates of return on projects on alternative modes.

The aim of this paper is to consider the main issues regarding appraisal methodology which were the subject of debate before the Leitch Committee, and to comment on the conclusions of that Committee. Firstly, we consider the context within which transport appraisal takes place. Then we discuss the basic technique of economic appraisal, as used in COBA. The next section of the paper considers the problem of bringing this together with the broader environmental aspects of appraisal. Following this, the problem of comparability between modes is discussed. Finally, we consider some problems specific to urban applications, which were outside the terms of reference of the Leitch Committee.

THE CONTEXT OF TRANSPORT APPRAISAL

Before discussing the techniques of transport appraisal themselves, it is essential to pay some attention to the context within which they are used. Transport appraisal may be used:-

(a) For strategic decision-taking—for instance, determining the level of the road construction programme, the size of the rail network etc.

(b) For determining whether a particular scheme is worth including in the programme as passing a cut-off rate of return, given the size and scope of the overall programme.

(c) For choosing between mutually exclusive schemes. (Schemes may be mutually exclusive on technical grounds—for instance, alternative routes for a single by-pass—or on financial grounds, given a budget constraint).

It is also necessary to take note of which body is undertaking the appraisal. This may be:-

(i) Central government, considering a scheme it will finance directly or by grants or loans (trunk roads, national rail investment).

(ii) Local government (local roads, restraint policies, bus services, rail services in the main conurbations).

(iii) Operators (British Rail, National Bus Company).

Firstly, regarding the type of appraisal at hand. It is generally the case that (c), choosing between mutually exclusive schemes is much easier than (b), determining whether a particular scheme is worthwhile, which is in turn easier than (a), the strategic level. For in decisions of type (c), the overall strategy is fixed, and one can focus on alternative ways of achieving that strategy. It is necessary only to consider differences in impact of the various schemes under consideration; their absolute value is not in question. Type (b) decisions are more difficult, requiring some judgement of the absolute worth of the project in question. Type (a) decisions are the most important and most difficult of all, and it is suspected that economic analysis play relatively little part in their determination, although the 1976 Consulation Document did state that, subject to various constraints, the achievement of reasonably equal rates of return at the margin in the trans-

port sectors for which the Department is responsible, remains a guiding principle [9].

Some illustrations here may be worthwhile. The kind of road appraisal that is made available at a public inquiry into a particular road scheme usually refers to the desirability of a particular relatively short stretch of new road *given* an overall rate of growth of car ownership and road traffic, *given* a strategy regarding investment in this and competing modes and usually *given* a decision (assumed or already taken) to build or to upgrade adjacent portions of the same route. Or to turn to rail investment, around 90% of which is dominated by the need to replace worn-out assets (although there is usually an element of 'betterment' about the new asset), there is often little room to manoeuvre once the strategic decisions about the future size of the network and the nature of services to be run have been taken. For instance, a particular resignalling scheme will usually include junctions of importance for one or more major trunk routes, and if these are to remain open it is usually the case that some action is essential. Appraisal, then, is confined to decisions of type (c) (e.g. number and location of signal-boxes, degree of track rationalisation to pursue).

The conclusion is, then, that whilst this sort of appraisal is undoubtedly necessary, the really important decisions have already been taken at a higher level. This is the reason for much of the discontent with the public inquiry system—from the point of view of the government, public inquiries are held to consider alternative ways of pursuing the same end (e.g. alternative routes for a new road), not to consider the absolute desirability of the new road and the merits of the strategy of which it forms a part. Indeed, until recently, discussion of the need for the new road was not even permitted at a public inquiry.

A similar issue arises regarding car ownership and traffic levels. It is not likely that the decision on an individual road scheme would have much influence on these variables. But recent evidence does suggest that government policy overall may play an important part. Thus, whilst the official forecasts of car ownerships in the past have assumed that the entire nation is progressing over time along a logistic curve towards a single saturation level of cars per head, with income and the real cost of motoring playing a minor part [24] new evidence ascribes the rise in car ownership to two variables, increases in real income and reductions in the cost of motoring [2].

Clearly, the latter is susceptible to government action (much of the reduction in the real cost of motoring in the 1960's was due to the reduction in purchase tax on new cars). Since the elasticities are found to be equal, it appears that the effect of a rise in real income could be entirely offset by an equal rise in the real cost of motoring. Moreover, car ownership at a given level of income varies greatly between locations, according to such factors as population density and the quality of public transport available [11, 13]. Again, planning and transport decisions appear to have a potentially significant role to play. If this is true of car ownership, it is likely to be even more true of car use. Yet, official forecasts have in the past simply assumed that the additional cars (even when second cars) will run the same average mileage per year as existing cars.

The second issue we raised in this section was that of the body for which the appraisal is undertaken. If for Central government, one would hope that it would be a comprehensive cost–benefit analysis, examining affects on all parties (although one would in general hope, in vain, for effects other than those on road users and the immediately affected population to be studied—for instance, effects on public transport operators). If for local government or for an operator, it would be unreasonable to expect the appraisal to take this form. Local government is answerable to its ratepayers, not to the population of the country as a whole. It therefore naturally tends to put their interests first. There are many examples of the implications of this for transport appraisal. For instance, a scheme which benefits travellers from elsewhere at the expense of the local population is in general unlikely to prove popular. On the other hand, attraction of shoppers, office workers and jobs from elsewhere is likely to appeal to the local authority both for reasons of rate revenue and—in times of general unemployment in particular—in preventing the social problems of unemployment. Such factors may appear to the local authority as benefits of improved transport, even if the effect is solely to relocate the facilities from elsewhere (and possibly from more needy locations).

Vol. 7
1979

For the transport operator, the range of costs and benefits for consideration is different again. At the forefront of his decision-taking is the financial objective he is set. Subject to achievement of this, he may have other objectives (for instance, maximisation of traffic in terms of passenger miles has been formally set as an objective for London Transport, and is undoubtedly influential with other transport operators). But no transport operator conducts his business with reference to the results of comprehensive cost–benefit analyses, (although these may be conducted on certain individual projects, particularly where application for government grants is involved).

THE TECHNIQUE

The basic comparison involved in the economic appraisal of transport in the trade-off of capital expenditure against time savings, accident savings and changes in vehicle operating costs. As will be seen from Table 1, on average 80% of the benefits of a scheme evaluated by COBA take the form of time savings, whilst 20% take that of accident costs. Changes in vehicle operating costs are not usually significant.

These are generally taken to be the 'hard' economic benefits of schemes. Nevertheless, there is still considerable debate about just how firm these benefits really are. Firstly, working time savings are valued at the wage rate plus labour-related overheads. The assumption necessary for this to represent the financial benefits to the employer is that the time savings are utilised either to increase production proportionately or to displace other labour. To what extent this does take place, as opposed to the time simply being taken as leisure, particularly for business car users who do not in any case work fixed hours, seems to be grossly under-researched. But even if the time is productively employed, for the financial savings to the firm to represent social benefits requires the further assumptions that labour is paid the value of its marginal social product and that any labour released is re-employed elsewhere.

TABLE 1. BENEFITS FROM A ROAD IMPROVEMENT SCHEME ON AVERAGE

Accident savings	20%
Vehicle operated cost savings	0%
Working time savings:	
Cars	26%
Light goods vehicles	11%
Heavy goods vehicles	11%
Buses	3%
Non-working time savings:	
Cars	23%
Buses	6%
Total	100%

Source [9], Vol. 2, Page 99.

Regarding non-working time savings, there is at least some empirical evidence that persons value journey time savings at approximately one-quarter of their wage rate, based on observations of behaviour where a choice exists between one mode of transport and another which is slower and cheaper [16]. Most of the case studies are based on the urban journey to work, however, and there is little evidence on how this valuation varies with journey purpose and trip length.

The third significant factor is accident cost savings. The bulk of the value imputed here reflects the loss of output of those injured and killed, which is valued at the wage rate. Again, this requires the assumption that wages are reasonable measures of marginal productivity. In any event, when an individual is killed, society loses his output, but at the same time has reduced demands placed upon it for consumption goods, investment and government spending. From the point of view of the rest of society, whether there is a net economic gain depends on the balance of these factors, which in turn depends greatly on the age and ability of the person killed. From the point of view of the person killed, there is of course a loss of consumption, but it seems somewhat extraordinary to treat this as part of the cost of being killed as is the case in the current procedure!

What is more important, however, is the value placed upon pain, grief and suffering and loss of life *per se*. There is no obvious market in which these can be valued, and one would doubt whether the average member of the public would see these as factors for which he could express the necessary monetary compensation. In practice, the decision on this valuation is largely political.

Thus we see that great doubt surrounds even those factors measured as 'economic effects' of schemes. No attempts have been made to place

confidence intervals on these valuations, but if they were it is hard to see the 95% confidence interval being less than ±50%. Nevertheless, whilst recognising the need for more evidence, the Leitch Committee was largely satisfied with this part of the appraisal procedure, and confined its recommendations to two significant changes. Firstly, pending further study, the value attached to pain, grief and suffering in the evaluation should be increased by 50%. Secondly, the common value of leisure time savings, the equity value of time, should be replaced by a value related to the income of the recipient, so that the value of savings to higher income groups and in wealthier parts of the country will receive proportionately more weight.

The first of these proposals appeared to express the widespread public concern that these factors were given too little weight in the appraisal. The second is more controversial. On the one hand, it makes valuation of this element consistent with valuation of the other elements in the appraisal, (previously, for instance, a low income person may be taken in the appraisal to prefer a fast, expensive service in circumstances in which he would really prefer a slow cheap one). On the other hand, it removed the only attempt to prevent the analysis giving more weight to a person's views, the richer he is. Had there been an attempt to allow for the distribution effects of transport investment more systematically in another way, the change would have been welcome. But the analysis by incidence groups recommended elsewhere in the Leitch report only partly does this, since the groups are not chosen specifically according to their incomes. As it is, the fear is that even less emphasis will be placed on distributive effects in future. In view of the evidence that transport investment has systematically tended to redistribute income from poor to rich, this is disturbing [7].

ENVIRONMENTAL EFFECTS

The treatment of the environmental effects of road schemes is important in respect of all the levels of decision-taking referred to above. Firstly, the value placed on environmental costs and benefits may raise or lower the average rate of return on road schemes as a whole, and thus influence the case for a bigger or smaller roads programme. Secondly, it may change the ranking of schemes within a given budget constraint. For instance, a high weight to environmental benefits relative to time savings may lead to a programme comprised almost entirely of modest by-passes rather than major new trunk roads. Thirdly, it may change the ranking of alternative proposals for a particular road scheme. For instance, it is often possible to reduce the environmental impact of a new road by additional expenditure and/or more devious routeing. Distinction between these decision-levels is important, since proposals which are helpful at one stage of the decision process may not be helpful elsewhere.

The ideal solution, in the eyes of many economists, would be to measure and value environmental costs and benefits in the same way as time savings and accidents are measured and valued. Unfortunately, all attempts at placing money values derived from consumer preference on environmental factors to date have involved such unlikely assumptions and so much sheer guess work that one can have very little faith in them. (For an example of this, see [22, Ch. 10] which results were quoted in the Consultation Document of 1976). On the other hand, these valuations could come from the political process, as is the case with the valuation attached to pain, grief and suffering. Presumably, the failure of such values to emerge is indicative of the much more complicated nature of the problems involved in analysing (and, indeed, measuring in physical terms) environmental effects.

The approach of the Leitch Committee to this problem is interesting. They set out a 'general framework' of around 80 relevant measures of the effects of transport schemes of which well over half could be termed broadly 'environmental'. A summary of these is given in Table 2. Comparison of alternatives is apparently to proceed by a process of pairwise comparisons. Now whilst this level of detail is very helpful in terms of taking detailed decisions at level (c) (choosing between alternative options for the same scheme) it is much less so at higher levels of the decision-taking process. This is recognised by the Committee in its statement that schemes should be given a priority rating of 1, 2 or 3 for inclusion in the programme. How the 80 or so dimensions of appraisal are to be reduced to a simple classifi-

Vol. 7
1979

TABLE 2. EXAMPLE OF THE LEITCH FRAMEWORK FOR THE ASSESSMENT OF TRUNK ROAD SCHEMES

	Incidence group	Nature of effect	No. of measures Financial	Other
1.	Road users directly affected	Accident savings	1	3
		Comfort and convenience		1
		Time savings	6	
		Vehicle operating cost savings	5	
		Amenity		2
2.	Non road users directly affected	Demolition or disamenity to owners of residential, commercial & industrial properties		22
		Demolition or disamenity to users of schools, churches, public open space		15
		Landtake, severance and disamenity to farmers		7
3.	Those concerned with the intrinsic value of the area	Landscape, scientific & historical value		3 (plus verbal discription)
4.	Those indirectly affected	Sterilisation of natural resources, land use planning effects, effects on other transport operators		6 (plus verbal description)
5.	Financial authority	Costs and financial benefits	7	
	Total		19	59

Source derived from [17, pp. 97–100].

cation into 3 groups, and how the results can be used in considering the overall size of the road programme is not spelt out. Clearly, fairly sophisticated multiple criterion decision techniques would be required since no-one could hope to weigh subjectively this number of items for a large number of schemes.

As soon as the need to consider such a wide range of non-financial measures is accepted, any justification for continuing to use arbitrary or political monetary valuations for any benefits loses force. The Leitch Committee went halfway towards recognizing this in that its 'framework' includes physical measures of accident victims even though the entire effects (including pain, grief and suffering) are supposedly included in the money valuations. But surely it would be worth reporting also physical volumes of time savings, disaggregated by trip purpose and journey length. These factors could then be weighed against accidents and environmental factors without the intervening mystique of money. Money values would be confined to items bought and sold in the market-place—i.e. primarily capital, maintenance and operating costs.

COMPARABILITY

The problem of comparability between the investment criteria used for different modes of transport arises at all three levels specified above. Firstly, alternative solutions involving different modes may be under consideration in the context of a specific problem. Here the solution postulated by the Leitch Committee is to subject the competing schemes to the same form of analysis, using the broader framework referred to in the previous section. Secondly, projects involving different modes may be competing for funds out of the same transport budget. Thirdly, broader strategic issues about the long run future of the transport system may be involved.

It is clear that economic appraisal can only be of use in the latter two decisions if the results for different modes are broadly comparable. Thus, even if some projects are subjected to financial appraisal for the internal decision-taking of a specific undertaking, these results need to be translated into a cost–benefit rate of return, and combined with other factors considered in the broader framework, if they are to be input into budgetary allocation and strategic decisions.

The two approaches may be expected to give the same answer only where the project at hand yields no costs or benefits to parties other than the operator concerned. The 1976 Consultation Document lists the following two cases where this may occur (their third example is

where, by coincidence, external benefits to one group are exactly offset by external costs to another):-

(1) Where there is no change in the price of the product and no rationing. The example given is that of a pure cost-saving investment. One should however, add two reservations. Firstly, product quality must not change. Secondly, if the investment changes marginal cost (as opposed to purely affecting overheads) then whatever the objectives of the organisation concerned, a price change is likely to result. Thus even pure cost-reducing investments often give rise indirectly to user benefits.

(2) Where the change in consumers surplus is negligible (e.g. because prices are raised to compensate exactly for service quality increase) *or* is captured by price discrimination.

Again this example seems to assume an unlikely behaviour pattern on the pattern of the operator. Even if he is seeking to maximise his profits, it is unlikely to be best to raise prices to a level which exactly offsets the quality improvement resulting from an investment project. Nor is it possible to recoup the majority of the benefits through price discrimination (except in freight transport, where negotiations with individual customers are the norm). Thus we must conclude that generally cost–benefit and financial returns on an individual project will differ. The Leitch committee examined a number of schemes which had been subjected to both methods of appraisal and found that in all cases examined the cost–benefit return was substantially above the financial return but with considerable variability in the ratio of the two. No single way of converting from one to the other could be found [15]. Indeed, it is possible for the financial return to exceed the cost–benefit return in special cases (e.g. where a project diverts traffic from another mode, but results in a cost saving that is less than the loss of revenue on that mode so that there is a net disbenefit to the competing operator to be taken into account).

The conclusion is, therefore, that strategic studies need to be undertaken to assess the cost–benefit rate of return on marginal schemes normally subjected to financial appraisal (such as improvements in inter-city rail services) before the Department could be in a position to pursue its stated objective of achieving reasonably equal rates of return at the margin in all the transport sectors. Even this leaves open the problems of assessing the wider considerations in the Leitch general framework.

CURRENT ISSUES IN URBAN TRANSPORT APPRAISAL

The Leitch report did not explicitly consider the problems of urban transport appraisal, since these were outside its terms of reference, although much of what it said applies here too. There are nevertheless some significant differences between urgan and inter-urban appraisals:-

(1) Network effects are relatively more important in the urban context, so that it is usual to evaluate alternative sets of projects for the entire area, rather than to consider schemes in isolation.

(2) Interdependence between modes is also usually stronger in urban than inter-urban markets, and the case—in terms of congestion and environmental effects—for deliberate action to direct traffic between modes is stronger.

For both of these reasons, concentration on the strategic level and attention to inter-modal effects has been stronger in the urban than in the inter-urban sector. Many studies have explicitly examined strategies based on rail or express bus services as well as the private car, and have often concluded that public transport based strategies are more desirable overall, even though they involve operating at a financial deficit. Once again, the 1976 Consultation Document suggested a considerable change in emphasis. The case for public transport subsidies in general was questioned, their justification being seen as solely on social grounds (to preserve a basic level of service and to help those for whom commercial fares cause hardship). The implication is presumably that public transport orientated strategies could only be pursued subject to their being financially viable, which would pose a major constraint

Vol. 7
1979

on the alternatives to be evaluated. On the other hand, the likelihood of public transport alternatives being viable could be increased by the simultaneous tightening of parking controls and other traffic restraint measures, and action to raise the cost of motoring.

One of the results of this approach is to stress, correctly, the inter-relationship between pricing policies, traffic restraint and investment appraisal. The measured benefit of a particular investment to relieve congestion will generally be higher, the less use has been made of alternative management techniques to deal with the problem. To this extent, investment and systems management techniques are partial substitutes for each other. However, it also should be borne in mind that some of the benefits of a public transport system may be dissipated by inappropriate pricing (e.g. Beesley and Foster, in [4], suggest that the benefits of the construction of the Victorian Line were substantially reduced by subsequent pricing decisions). The Consultation Document apparently discounts the idea that public transport prices can influence modal split. It also fails to appreciate that public transport pricing decisions may be evaluated in much the same way as investment decisions, and that in such an evaluation a large part of the financial effect of a change in fares represents a transfer between parties, rather than a net cost or benefit. For instance, a study of the consequences of making public transport free in the city of Boston concluded that this would involve a financial disbenefit of around $75 m [17]. Of this, some $70 m consisted of lost revenue (a transfer payment) and only $8.6 m was due to the cost of carrying additional traffic. Even this would be partly offset by a saving of $3.5 m in fare collection costs. Thus the real resource cost of the policy would be around $5 m, and against this must be set the benefits of increased public transport use. Again, there is clearly a big difference between the results of cost–benefit and financial appraisals.

CONCLUSION

This survey of the use of economic appraisal in transportation analysis has, of necessity been extremely limited in scope. It has concentrated on recent debates in the UK, as presented in the Consultation Document and the Leitch Report. Even so, it has not been possible to be comprehensive, much of the valuable work of the latter committee, for instance on sensitivity testing and international comparisons, has been ignored. Nevertheless, it is hoped that a sufficient case has been made for the following three conclusions:-

(1) Economic appraisal of transport schemes can only be a part of a much wider appraisal. Given this, it is misleading to isolate certain non-market effects and attach monetary valuations of dubious reliability to them whilst leaving the majority of effects in physical terms.

(2) It is much easier to take these broader non-economic effects into consideration at the level of comparison of individual schemes than at the strategic level. Even so, there is no justification for an approach whereby strategic decisions are taken on purely quantitative evidence and environmental factors only taken into account in choosing between mutually exclusive schemes.

(3) The most efficient transport policy will be one that allocates any given transport budget according to a common criterion throughout all the sectors of the transport industry. Even if this is not regarded as feasible, common criteria should be used in choosing between interdependent or mutually exclusive projects from different sectors of the industry, and in taking strategic decisions on the allocation of funds between sectors.

Such an approach to decision-taking is within the traditional framework of economic appraisal, in stressing identification of costs and benefits to interested parties. But it involves a much broader consideration of costs and benefits, and a greatly increased emphasis on qualitative factors, than has been usual in transport appraisal in the past. Techniques for handling this sort of information on a large scale, despite the growing interest in goals achievement matrix formulations and multiple criterion decision techniques, [19], are still very much underdeveloped.

REFERENCES

1. ADAMS JGU (1974) Saturation planning. *Tn & Ctry Plann.* **42**(12), 550–554.
2. BATES J, GUNN H & ROBERTS M (1978) *A Disaggregated Model of Household Car Ownership.* Dept. of Transp. Res. Rep. **20.** Dept. of Transp., London, UK.
3. BEESLEY ME, COBURN TM & REYNOLDS DJ (1960) *The London–Birmingham Motorway—Traffic and Economics.* Road Res. Lab. Tech. Pap. **46.** Crowthorne, Berks, UK.
4. BEESLEY ME & FOSTER CD (1965) The Victorian Line: social benefits and finances. *Jl R. Stat. Soc.* Ser. A, **128**, 67–88.
5. Cmnd **3686** (1968) *Transport in London.* HMSO, London, UK.
6. COMMISSION on the Third London Airport (1971) *Report on the Siting of the Third London Airport.* HMSO, London, UK.
7. DALVI MQ & NASH CA (1977) The redistributive impact of road investment. In *Urban Transportation Planning* (Eds BONSALL P, DALVI MQ & HILLS PJ) pp. 197–217, Abacus Press, Tunbridge Wells, UK.
8. DAWSON R (1968) *Assessment of Road Improvement Schemes.* Road Res. Lab. Tech. Pap. **75.** Crowthorne, Berks, UK.
9. DEPARTMENT OF THE ENVIRONMENT (1976) *Transport Policy: a Consultation Document.* HMSO, London, UK.
10. DUPUIT J (1844) On the measurement of the utility of public works. *Annls Ponts Chauss.* 2nd Ser., **8**, [Translated into English BARBACK RH (1952)] *Int. Econ. Pap.* **2**, 83–110.
11. FAIRHURST MH (1975) The influence of public transport on car ownership in London. *Jl Trans. Ecs. & Pol.* **9**(3), 193–208.
12. FOSTER CD & BEESLEY ME (1963) Estimating the social benefit of constructing an underground railway in London. *Jl R. Stat. Soc.* Ser. A, **126**, 46–58.
13. FOWKES AS, BUTTON KJ & PEARMAN AD (1978) *The sensitivity of car ownership of households in West Yorkshire to household accessibility.* PTRC An. Summer Meet., London, UK.
14. GWILLIAM KM & NASH CA (1972) The evaluation of urban road schemes: a comment. *Appl. Econ.* **4**, 307–315.
15. HARRISON AJ & MACKIE PJ (1973) *The Comparability of Cost–Benefit and Financial Rates of Return.* Govt Econ. Serv. Occ. Pap. **5**, HMSO, London, UK.
16. HARRISON AJ & QUARMBY DA (1969) The value of time in transport planning: a review. ECMT 6th Round Table, Paris, France. Reprinted in *Cost–Benefit Analysis* [Ed. LAYARD R (1972)] Penguin, Harmonsworth, Middx, UK. 173–208.
17. KRAFT G & DOMENLICH TA (1972) Free transit. In *Reading in Urban Economics* (Eds EDEL M & ROTHENBURG J) pp. 459–480, Macmillan, New York, USA.
19. LICHFIELD M., KETTLE P. & WHITBREAD M (1975) *Evaluation in the Planning Process.* Pergamon Press, Oxford, UK.
20. MISHAN EJ (1970) What is wrong with Roskill? *Jl Trans. Ecs & Pol.* **4**(3), 221–234.
21. PLOWDEN S (1967) Transportation studies examined. *Jl Trans. Ecs & Pol.* **1**(1), 5–27.
22. PRYKE R & DODGSON J (1975) *The Rail Problem.* Martin Robertson, London, UK.
23. SELF P (1970) Nonsense on stilts: the futility of Roskill. *Polit. Q.*, **41**, 249–260.
24. TANNER JC (1974) *Forecasts of Vehicles and Traffic in Great Britain.* Transp. & Road Res. Lab. **LR650**, Crowthorne, Berks, UK.

ADDRESS FOR CORRESPONDENCE: *CA Nash, Esq, Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK.*

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 5, pp. 451 to 458
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1001-0451$02.00/0

Programme Budgeting in the Health Sector

ALASTAIR GRAY

ROBERT STEELE

University of Aberdeen, Scotland

(*Received May* 1979)

The growth of the public sector has been accompanied by an increased interest in this sector among economists and non-economists alike. Much of this interest has been directed towards the overall size of the public sector, and the efficient allocation of resources *within* the public sector has remained comparatively neglected. This paper outlines some of the problems of resource allocation facing decision-makers in the public sector, and reviews the actual and potential uses of one decision-making tool—programme budgeting—in improving resource allocation. The health care sector is chosen to illustrate the technique of programme budgeting, the practical and conceptual problems of implementation, and the limitations as well as the possibilities offered by this approach.

INTRODUCTION

BETWEEN 1960 and 1975 expenditure in the Public Sector in the UK grew from 41% of Gross National Product to 58%. In absolute terms this represented an increase from £9,451m to £54,465m. Expenditure on the National Health Service (NHS) has risen over the same period from 9.5% of total public expenditure to 9.6% [13].

In terms of sheer size, therefore, expenditure on the NHS is an important part of public expenditure, and the management of the NHS budget equally important in the context of general management and planning in the Public Sector. This paper is not concerned with the allocation of resources between different parts of the Public Sector (health, social services, defence, etc.) but their allocation *within* the health service. A *planned* allocative framework for resources is necessary both in the health service and in many other areas of the public sector because of the non-existence of the interplay of economic forces which predominate in the private sector.

A major characteristic of much of the public sector is the absence of a market price mechanism. Education, social services as well as health care, are largely zero-priced at the point of consumption. Thus the level of provision of a particular good or service must be determined without reference to either a conventional demand or supply schedule; to a large extent demand is supply-determined and rationing occurs more through the political process than through the price mechanism.

Indeed in the NHS, the concept of demand tends to be replaced by that of 'need'. Need has been defined in several ways [2, 5, 27]. Almost always however, such definitions embrace the idea of some third party being involved in the valuation process, thereby at least partially removing the more normal assumption of consumer sovereignty which lies behind the concept of demand. For example, Cooper [6] suggests that: "'Needs' are those demands which in the opinion of the doctor require medical attention. That is, they are an expert's view of our health state". This need, as Williams [38] indicates, is based on the concept of merit goods (i.e. those goods such as compulsory education for children where it is

assumed that experts are better placed than individual consumers to decide what is in the best interests of the individual consumers).

This is not the place to debate the pros and cons of the concept of need. Rather the intention is simply, first, to indicate its presence in health service planning discussions and secondly to draw attention to the potential breakdown in health care of the more normal concept of demand. Insofar as consumer sovereignty does not apply in the NHS then it becomes necessary to provide a framework within which judgments by experts (more commonly the medical profession) can be made on a rational, explicit basis.

The pervasiveness of zero-prices in the NHS also creates problems on the supply side. Normally (i.e. in market situations) prices rise when there is shortage which results in more resources being attracted into the area. Where prices are absent it becomes possible to define shortage only in terms of some, almost inevitably, arbitrary standard of what is required or needed. Further on the supply side, the fact that so little remuneration of medical practitioners is based on fees for services, results in the possible separation of wage rates from any productivity measures.

Thus on both the demand and supply sides zero-pricing at the point of consumption and the methods of remuneration in the NHS mean that the normal rationing mechanism of the market is missing or is at best muted. This is not to argue that health services should not be provided in the public sector but simply to illustrate the need for other mechanisms for determining how best to allocate scarce NHS resources. (For a discussion of the 'public versus private' debate in health care see, for example [8, 21, 22, 33].)

Without the decision-making format provided by normal market forces some other framework needs to be developed to aid explicit and coherent policy-making. Of course, even although it could be argued that such a framework has not existed in the NHS and that the fundamental aim of the NHS of 'the provision of the best medical care for all' denies the need for choice, there are inevitably numerous decisions being made involving economic concepts of rationing, trade-offs, and implied valuations. Unfortunately many of these decisions tend to be made in an '*ad hoc*' fashion with no *explicit* acceptance of either choice or valuation. Planning then becomes a response to past decisions and choices instead of decisions and choices being a consequence of planning. Because of this, inconsistencies, inefficiencies and inequalities prevail. It is in this context that this paper discusses the use of programme budgeting as a technique to assist in improving the allocation of resources in the National Health Service.

HISTORY OF PROGRAMME BUDGETING

Programme budgeting is not a new technique in public sector decision-making. It appeared in its earliest form in 1961 in the US Defence Department, having been developed by analysts working within the RAND Corporation [17, 28]. The reason for choosing the Defence Department lay in the magnitude of its expenditure budget, unease over public control of this budget, and the enthusiasm of Secretary of Defence McNamara for budgeting reform and control. After a period of experimentation in this Department, President Johnson recommended the comprehensive introduction of the technique for all Federal Government Departments in 1965.

The need for reform in many of these departments was striking. For instance, the Department of Health Education and Welfare had overall responsibility for three important areas of government activity, and yet in practice had no unified organisational or budgetary system. In 1965 42 departments, agencies and bureaux were involved in the finance of education alone. This made it virtually impossible to estimate the character, importance, and effectiveness of government spending on education. The same held true of health care, and as Glennerster [15] points out, from this position "merely to consider all health expenditure together would be a considerable advance". Thus, by 1967 the health programme had been split into 12 sub-programme headings, and expenditure from all government sources grouped together under these headings [23].

However, it became increasingly apparent that the Federal Government was placing unrealistically high hopes on the benefits programme budgeting could offer—President Johnson referred to 'identifying national goals

Vol. 7
1979

with precision' and attaining them in the most cost-effective way possible. The practical difficulties of measurement of the aims, objectives and output of public sector departments were neglected, and the inevitable failure to realise expectations led to an understandable decline in enthusiasm for programme budgeting in the USA. (For a critical analysis of American experience of programme budgeting, see [35, 36].)

Given these false hopes and failed expectations, we should perhaps look to the USA for mistakes to avoid rather than for indications of the practical usefulness of programme budgeting.

In contrast to the US experience, programme budgeting in the UK began modestly and has built on its successes. It made its British debut in the Department of Education in 1968 [9], and thereafter began to permeate other central government departments. For instance, in 1970 the Home Office developed a programme approach to police expenditure [34] to assist planning in individual police forces. It regrouped the traditional accounts-based expenditure—pay, allowances, premises, pensions, etc.—into seven major programmes, relating expenditure to programmes such as ground cover, crime investigation and control, traffic control, etc.

1970 also saw the introduction of programme budgeting into the Department of Health and Social Security (DHSS). The stimulus for its introduction there was a simple one. When poor conditions in a Cardiff hospital for mentally handicapped children were revealed in the press, the Department undertook to increase expenditure on this client group, only to discover that nobody knew how much was actually spent on care of the mentally handicapped! Moreover it did not know if spending had been increasing or decreasing, though it did possess accurate figures on, for instance, expenditure on split peas by hospitals. There was clearly a need to regroup expenditure information under programme headings.

Further stimulus to planning and particularly to programme budgeting came with the reorganisation of the health service in 1974. Prior to this date the hospital and primary care services, although funded centrally under the NHS, were separate entities, while the community/domiciliary services were the responsibility of and funded by local government. Certainly in terms of *patient* management the tripartite services interacted and liaised but the situation was not conducive to broad resource management and planning. This situation was rectified in 1974 when the community services were embraced by the NHS and their funding switched from local to central government. A new administrative structure was created by Regional, Area and District Authorities in England (and Boards in Scotland); the need for greater coordination and consultation was explicitly recognised; and the foundations laid for increased scope for improved management and planning. (For more details of NHS organisation and reorganisation see [24].)

The acceptance of programme budgeting in DHSS is now reflected in the administrative reorganisation of the Department around meaningful programme (i.e. patient care groups such as the elderly) planning groups. The publication of the priorities documents in England [10, 12] reflected the need for guidance to the newly created health authorities on planning, particularly in the context of the relative priorities to be given to different care groups, such as the mentally ill. These documents also indicated the central role which programme budgeting was playing in creating a framework for assisting in deciding these priorities. In turn it then became desirable for planners and decision-makers beneath central government level to adopt a similar planning approach and a number of Health Authorities (in England) and Health Boards (in Scotland) are gradually beginning to think in terms of introducing programme budgeting into their management accountancy systems.

THE NATIONAL HEALTH SERVICE: SOME BACKGROUND THOUGHTS

Questions of choice and valuation exist throughout the public sector, and the public provision of health care is no exception. What is perhaps exceptional is the reluctance of those who influence resource allocation within the NHS to accept that these choices have been made in the past, are being made at present, and will have to be made in the future.

The origins of this attitude are deeply embedded in the ethical and emotional issues raised by health care. When the NHS was established the central object was to provide

the *best* medical care possible for *all*. Allied to this was the belief that the quality and comprehensiveness of the all caring services supplied through the NHS would quickly erode the pre-existing stock of ill-health, and by increasing the general level of health would reduce demand for health care and hence expenditure on health care.

It is now well known history that events did not follow this course. Yet despite clear evidence that the NHS would—and did—grow at a relatively rapid pace through both the fifties and the sixties, the initial view of it in terms of providing the best at falling cost resulted in an unfortunate legacy, based on three key factors:

(i) the apparent lack of willingness to accept that choices not only had to be made but were being made;

(ii) the continuation of the view that the place to be was always on the technological frontier of medicine; and

(iii) the constant chasing of the 'best' in medical care.

This legacy was simply a reluctance to make rational choices in health care resource allocation and the almost total lack of explicit recognition of the concept of opportunity cost i.e. that pursuing one particular policy meant an opportunity forgone in failing to achieve some other aim. This legacy more than any other is responsible for most of the resource allocation problems now built into the NHS.

For once the concept of opportunity cost is accepted there is nothing which is absolute, everything becomes relative. Different objectives require differential relative values. Relative weights have to be attached to different outputs.

As discussed above, the concept of need pervades much of the discussion of planning in the NHS. The reluctance on the part of NHS managers (and particularly the medical managers) to accept not only that resources were scarce but also that need was not finite led to the view that all need not only could be met but should be met. Against such a background it is not surprising that issues of explicit choice in resource allocation in the NHS and the related need for planning were slow in being recognised. Who needs planning if needs are thought to be finite and resources infinite?

In practice the service has tended to operate under what can only be described as a form of 'disjointed incrementalism' [25]—*ad hoc* micro-level decision-making with little or no overall planning. Certainly some part of the reason for this lies in the diffuse decision-making structure of the NHS and the interactions which occur in the decision-making process between different levels and different groups. Decisions are made and influenced by central government, regional authorities, areas, districts, sectors and units; by consultative and advisory committees at national and local level; by individual clinicians; by local health councils and the public at large. This means first that there is no single level or group at which the ideas of planning should be aimed and second that problems arise in integrating the roles of the different decision-makers in any planning process.

Not only are the decision-makers or 'choosers' numerous; so also are the choices. There are sectoral choices to be made both within and between hospital, community, and primary care services. There are issues of choice related to the different final outputs of health care—chronic versus acute, preventive versus curative; and there are choices between client groups: elderly, children, mental handicapped, etc. and between disease groups: cancer, kidney disease, etc. Finally there are choices related to equity considerations—equal access for different socio-economic groups, the location of facilities geographically, etc.

Given these different levels and natures of choice and given the many different groups of decision-makers, it becomes clear that some form of planning framework is required in which to set at least some of these issues of choice. It is on programme budgeting that much of the onus has been placed for improving planning in the NHS.

PROGRAMME BUDGETING IN THE NATIONAL HEALTH SERVICE

The principle underlying programme budgeting in the NHS is very simple. If decisions are to be made about broadly defined health

Vol. 7
1979

care objectives and priorities then data need to be provided in similarly broad terms to match the nature of these choices. The simplicity of the concept belies the practical problem inherent in health service choices. For instance what are the objectives of care of the elderly? What relative priorities are to be attached to the treatment of cancer compared with the prevention of heart disease? Ideally we would want to relate the resources involved in different programmes to final outputs of programmes in terms of clearly defined objectives, whether they be reduced mortality, morbidity, or increased comfort and care. But while there is an expanding interest in several disciplines in the measurement of health output (see for example [7, 14]) measurement of health—indeed obtaining an operational definition of health—remains a very real problem in any evaluation of health services. Even the output of reduced mortality raises problems of measurement when aggregation has to be made of lives saved across, for example, different age groups. Changes in health status, when related to morbidity, are difficult to quantify, particularly when different diseases and severities of illness have to be compared. It is perhaps not surprising that two despairing authors (Lave and Seskin [20]), have written that: "The most unequivocal measure of health status is death"—and even then, as indicated, lives, and consequently deaths, may differ!

Added to this difficulty is the fact that so little is known about the production functions of health care—i.e. the relationship between factors of input to a particular care regime (medical and nursing manpower, support services, laboratory tests, etc.) and the outputs in terms of improved health (but see, for example, [19, 32]).

The reasons for this are several but perhaps the main one is that only in the last decade have randomised controlled trials become accepted [4]. These trials offer one of the few ways of assessing, in quantified terms, what the impact or effect of a particular treatment regime is. But the majority of treatment regimes which comprise health services have not been subject to such studies and their *quantified* impact is therefore unknown.

To circumvent these problems it has been necessary to use 'intermediate' measures of output. These usually take the form of utilisation or activity data such as discharges, bed utilisation, and visits by community services and attendances at clinics or surgeries. While these are perhaps related as much to inputs as to outputs, progress can be made with them, particularly if the issues of effectiveness are always recognised and can be subjectively judged by experts. This combination can provide a sufficient base on which to build the output side of the programme budget.

A few points about the nature of programme budgeting are worth emphasising.

First, since programme budgeting is intended to provide data relevant for 'broad brush' planning it is not necessary to present precise details in the budgets. Programme budgeting is a planning framework which can influence the decision-makers not only to take an overview of the health services generally, but to take an overview which relates to broad health care goals [30]. It is not an evaluative technique *per se*.

Secondly, the structure of the budgets should be output orientated in terms of realistic, even if not quantifiable, health care objectives. There is no single correct way of structuring the budgets; indeed, the classification should vary according to the nature of the trade-offs to be made. The most likely choices in health care are (i) client groups (the elderly, children, mentally ill, etc.); (ii) disease groups (digestive system diseases, neoplasms, etc.); and (iii) geographical groupings (for example, the RAWP Report [11] was concerned with equity between geographical regions and the 'programme structure' in this instance was therefore regionally based). Again, if programme budgeting is used within an individual programme then appropriate sub-programmes can be devised (for example, for maternity these could be ante-natal, labour/birth and post-natal).

Thirdly, the presentation of the programme budgeting data takes three forms for three related purposes:

(a) The current deployment of resources across the programme heads. This provides a 'snapshot' of how resources are presently being devoted to the different groups and hence to different objectives.

(b) Past trends up to the present position. These indicate how the programmes got to

where they are, the growth paths of resources and outputs and, by deduction, the extent to which policies in the past have affected development. Thus, the study of trends gives an opportunity for monitoring and, once the programme budgeting approach is established, a means of checking fairly quickly whether what was intended has been occurring.

(c) Lastly, future planning. Various scenarios related to the different programmes can be presented, indicating within the expected constraints of future budgeting allocations what possible policies might be pursued for and between different programmes and what the resource and output impacts of these are likely to be.

PROBLEMS OF IMPLEMENTATION OF PROGRAMME BUDGETING

The current accounting system in the Health Service is very much related to inputs, i.e. nursing salaries, X-ray services, heating and lighting, etc. Indeed the accounting and coding systems in the NHS are largely designed for the purposes of monitoring input use and controlling expenditure. They are not designed for planning purposes.

This means that in the context of programme budgeting either a new system of accounting is required or the existing system has to be adapted to allow costs to be allocated to programme heads. The first of these could well prove to be extremely expensive; and the second inevitably means that some precision is lost. Against this background it is perhaps not surprising that few Health Authorities have embarked on serious exercises in programme budgeting.

A second difficulty lies in the fact that the output measures available for programme budgeting are less than perfect. Ideally one would want to measure the outputs of health services in terms of lives saved, reduction in morbidity, increased comfort and care, etc. However as indicated previously the paucity of available data on health status and on production functions means that in the great majority of health care regimes precise measurement of changes in health status is not possible and the necessary information on these production functions is not available. Consequently for the present, reliance has to be placed on intermediate outputs such as numbers of patients discharged, numbers of nurse visits in the community, and the effectiveness of different regimes left to the judgement of clinicians. The deficiencies in these approaches are obvious. Here we have a further reason why health care planners may have been slow to use programme budgeting for planning purposes.

A further difficulty lies in the fact that decision-making in health care is so diffuse both in terms of the levels at which different types of decisions are made and also in terms of the number of different individuals involved in decision-making on resource allocation. It is perhaps inevitable, against this background that individual decision-makers can lose sight of the broad objectives and the priorities attached to these of health care in general.

Finally, and perhaps sadly, it has to be admitted that the NHS has not shown itself particularly willing or enthusiastic about planning techniques. This is perhaps a function of the fact that the service has been able to expand quite considerably from year to year and consequently the issue of choice has been seen if it has been seen at all in terms of simply delaying the implementation of certain desirable policies. If the issues of choice, trade-offs, and opportunity costs are not seen, then it is perhaps inevitable that the potential value of programme budgeting as a planning framework is understated in the minds of policy makers in the NHS.

Vol. 7
1979

EXAMPLES OF PROGRAMME BUDGETING IN THE NHS

The most noteworthy example of programme budgeting in the NHS is that undertaken by the DHSS and which formed the framework for the priorities documents of the 1970's (DHSS [10, 12]).

The 1976 consultative document outlined a strategy for health and local authorites in formulating their own regional and area plans up to 1979/80. The programmes selected were a reflection of the fact that both health and local authorities were involved, in that a strict 'client group' criterion was avoided in favour of a grouping of services by major users. Thus, as

the document points out [10, p. 78], all home nursing was included in the 'elderly' programme, and all health visiting under 'children', although both home nurses and health visitors serve a range of clients. This approach underlines the 'broad-brush' nature of programme budgeting.

For hospital services, expenditure was split into in-patients and out-patients and day patients; thereafter unit costs were derived per bed week and attendance. These unit costs were then applied to all bed weeks or attendances within the specialty grouping to give an estimated total expenditure figure. As the document points out, although this allocation is not precise, estimates for future years are not very sensitive to base year allocation errors. The only aggregate effect is that a misallocated sum may have a different growth rate applied to it.

Using growth trends within programmes from 1970/71–1973/74 it was possible to project from current expenditure levels the likely outcome in 1979/80 if spending patterns remained unchanged, and to compare this projection with the *desired* outcome in 1979/80. Juxtaposing this with data on changing population structure, utilisation rates and activity rates, a clear picture of trends within each programme emerged. For instance, the cost of hospital maternity services had risen by about 4% per annum between 1970 and 1973, yet the total number of births, of in-patient cases, of out-patient attendances, and of beds all fell. This suggested that the momentum of past expenditure increases had carried on into a period of declining utilisation, with consequent increases in cost per case. By way of contrast, population changes were placing an increasing strain on services for the elderly, which could not respond without a higher rate of growth of expenditure.

Two points should be drawn from this: first, these results do not in themselves suggest that 'too much' was being spent on maternity services; it may have been an implicit objective of planners to raise standards in maternity care by maintaining expenditure despite falling utilisation. The information presented simply facilitated decisions about priorities. Secondly, it is clear that priorities must respond to demographic, social, and epidemiological change, and therefore that a programme budgeting exercise must be continually updated and redefined.

At a local level, a programme budgeting structure was created for one health board in Scotland in a very different way [29] and is currently the subject of experimentation in most health boards in Scotland. The method used in this case was simply to attribute past expenditures to different programmes (again client groups) directly, where this was possible (e.g. doctor and nurse salaries) and indirectly by apportionment on, for example, an in-patient *per diem* basis for items such as catering, where it was not. Again, as in the DHSS format, outputs were measured by discharges, visits, etc. From past trends in expenditure and outputs it was then possible to extrapolate these to create a framework for planning future resource allocation.

At a somewhat different but related level a number of studies have been made of disease and specialty costs [1, 26, 31]. Although these have a different purpose, nonetheless in terms of improving the methodology for devising programme budgets, they are relevant to any discussion of the state of the art of programme budgeting in the NHS.

CONCLUSION

It will be clear from what has been said that programme budgeting in the health service should be seen as simply a planning framework within which broad health objectives and priorities can be discussed. It is not in itself evaluative. For evaluation it is necessary to turn to the techniques of cost-effectiveness and cost-benefit analysis. (See for example, [3, 16, 18, 37].)

It will also be clear that while programme budgeting can be a useful planning tool in health care, the extent to which it is currently used in the NHS is limited. Partly this reflects the lack of appreciation of the need for rational planning in the NHS; but partly it reflects the problems involved in implementing it, both in terms of the existing accounting system in the NHS and in terms of the inherent problems surrounding the nature of health and health care. Certainly there is a growing awareness in the NHS of the necessity for more synoptic planning and this has been explicitly recognised in the reorganised health service in Eng-

land. There are also signs that more thought is being given to designing accounting systems which will be more appropriate for planning purposes. Consequently, while programme budgeting remains at present underutilised in the NHS, there are grounds for believing that the next few years will see its being used much more widely.

ACKNOWLEDGEMENTS

The authors are grateful for the advice and assistance of colleagues within HERU, particularly Ann Chamberlain and Gavin Mooney.

REFERENCES

1. Babson J (1973) *Disease costing.* Manchester University Press. Manchester, UK.
2. Bradshaw J (1972) A taxonomy of social need. In *Problems and Progress in Medical Care: Essays on Current Research* (Ed. McLachlan G) 7th series. Oxford University Press, London, UK.
3. Bunker JP *et al* (Eds) (1977) *Costs, Risks & Benefits of Surgery.* Oxford University Press. New York, USA.
4. Cochrane AL (1972) *Effectiveness and Efficiency: Random Reflections on Health Services.* Nuffield Provincial Hospitals Trust, London, UK.
5. Cooper MH (1975) *Rationing Health Care.* Croom Helm, London, UK.
6. Cooper MH (1974) Economics of need; the experience of the British NHS. In *The Economics of Health and Medical care* (Ed. Perlman M). MacMillan, London, UK.
7. Culyer AJ, Lavers RJ & Williams A (1971) *Social Indicators: Health.* Social Trends **1**(2). HMSO, London, UK.
8. Culyer AJ (1972) On the relative efficiency of the NHS. *Kyklos* **25**(2) 266–287.
9. Department of Education & Science (1970) *Output Budgeting for the Department of Education and Science.* HMSO, London, UK.
10. Department of Health & Social Security (1976) *Priorities for Health and Personal Social Services in England: a consultative document.* HMSO, London, UK.
11. Department of Health & Social Security (1976) *Sharing resources for health in England.* Report of the Resource Allocation Working Party, HMSO, London, UK.
12. Department of Health & Social Security (1977) *Priorities in the Health and Social Services: The Way Forward.* HMSO, London, UK.
13. Elliott RF & Fallick JL (1979—Forthcoming) *Pay in the Public Sector.* MacMillan, London, UK.
14. Fansehl S & Bush JW (1970) A health status index and its application to health services outcome. *Ops. Res.* **18**(5), 1021–1066.
15. Glennerster H (1975) *Social Service Budgets and Social Policy: British and American Experience.* Allen and Unwin, London, UK.
16. Hagard S, Carter F & Milne RG (1976) Screening for spina bifida systica: a cost–benefit analysis. *Br. Jl Prev. & Social Medicine* **30**(1) 40–53.
17. Hitch CJ & McKean RM (1960) *The Economics of Defence in the Nuclear Age.* Harvard University Press, Cambridge, Massachusetts, USA.
18. Klarman HE (1967) Present status of CBA in the health field. *Am. Jl Publ. Hlth* 1948–53.
19. Klarman HE *et al* (1968) Cost-effectiveness analysis applied to the treatment of chronic renal disease. *Med Care* **6**, 48–54.
20. Lave LB & Seskin E (1977) *Air Pollution and Human Health.* Johns Hopkins Press, Baltimore, USA.
21. Lees DS *et al* (1964) Monopoly or Choice in Health Services? London, I.E.A.
22. Lees DS (1967) Efficiency in Government spending—social services: health. *Pub. Fin.* **22**(1/2), 176–189.
23. Levin A (1971) Multi-secting the nation's non-defence programs. *Publ. Admin. Rev.* **31**(2), 171–186.
24. Levitt R (1976) The Reorganised National Health Service. Croom Helm, London, UK.
25. Lindblom C (1959) The science of muddling through. *Publ. Admin. Rev.* **19**(2), 79–88.
26. Mason AMS, Perry JA & Skegg JL (1973) *Disease costing in hospitals: a review of completed work.* Introductory paper for a one-day symposium, Northwick Park Hospital. (Mimeo).
27. Matthew GK (1971) Measuring need and evaluating services. In *Portfolio for Health* (Ed. McLachlan G). Oxford University Press, Oxford, UK.
28. McKean RM (1958) *Efficiency in Government and the Use of Systems Analysis.* John Wiley, New York, USA.
29. Mooney GH (1977) Programme budgeting in an area health board. Hospital and Health Services Review.
30. Pole JD (1974) Programmes, priorities and budgets. British Journal of Preventive and Social Medicine **28**, 191–195.
31. Russell EM *et al* (1974) Patient costing study. Scottish Health Services Studies. No. 31. Edinburgh. Scottish Home and Health Department.
32. Simpson PR, Chamberlain J & Gravelle HSE (1978) Choice of screening tests. Journal of Epidemiology and Community Health **32**(3), 166–170.
33. Titmuss RM (1963) *Essays on the Welfare State.* Unwin, London, UK.
34. Wasserman GJ (1970) Applying PPB to Police Expenditure, O and M Bulletin (The Journal of Management Services) **25**(4), 197–210.
35. Wildavsky A (1964) *The Politics of the Budgetary Process.* Little, Brown, Boston, Massachusetts, USA.
36. Wildavsky A (1968) The political economy of efficieny: cost–benefit analysis, systems analysis and programme budgeting. In *Planning Programming Budgeting: A Systems Approach to Management* (Eds Lyden FJ & Miller EG). Markham, Chicago, USA.
37. Williams A (1974a) The cost–benefit approach. *Br. Med. Bull.* 252–256.
38. Williams A (1974b) 'Need' as a demand concept (with special reference to health). In *Economic Policies and Social Goals: Aspects of Public choice* (Ed. Culyer AJ). Martin Robertson, London, UK.

ADDRESS FOR CORRESPONDENCE: *A. Gray, Esq, Health Economics Research Unit, Department of Community Medicine, University of Aberdeen, AB9 2ZD, Scotland, UK.*

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 5, pp. 459 to 467
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1001-0459$02.00/0

Cost-Effectiveness in Defence Expenditure

JOHN M TREDDENICK

University of Aberdeen, Scotland, UK

(*Received May* 1979)

Decision-making in defence is complicated by the need to make long-range plans with regard to weapons systems and force structures in an environment marked by great strategic and technological uncertainty. Cost-effectiveness analysis, which is a variant of the general analytical approach, is an attempt to clarify and structure defence choice situations. As a method of analysis, it shares certain properties of both cost–benefit and operational analysis, and indeed emerged from them in response to the unique needs of defence planners. Cost-effectiveness analysis involves a number of conceptual and practical problems, but its most difficult challenge is the requirement to exist in a political and bureaucratic environment where the system of incentives may not be compatible with the notion of economic rationality. However, the recent appearance of competitive cost-effectiveness analysis of defence options promises to promote the relevance and applicability of this approach.

INTRODUCTION

DEFENCE CHOICES are agonizingly difficult. In a world of limited resources, the sums required to maintain military forces impose very real burdens on national economies. Yet, in a world of perceived threats, actual confrontation and not infrequent outbreaks of armed conflict, defence expenditures seem a prudent investment. But what constitutes an appropriate level of such expenditures, and how should they be allocated?

The crucial difficulty is that one is never quite sure about the return for the expenditures. Outlays provide means, but the links between means and the ultimate ends of defence are obscure, if for no other reason than that, in the absence of outright war, the ends themselves are obscure. An enormous and pervasive element of uncertainty characterises the defence decision-making environment. It attaches first of all to the strategic context in which decisions will be implemented.

Decisions can be invalidated by unforeseen initiatives of potential opponents, by unexpected reactions of those opponents to one's own initiatives, and by the actions and reactions of allies and non-aligned nations. Uncertainty also attaches to the technological context. Military forces operate at the frontier of technology, which advances at a dizzying rate and in unexpected directions. In contrast to the rate of strategic and technological change, the development and production of advanced weapons is a time-consuming process. It may take ten years or more for systems to go from concept to operational status; in that time they may very well have become inappropriate to changed strategic circumstances or obsolete relative to the state of the art. At the same time, a system's service life may be 20 or 25 years; and, because of high and rising replacement costs, there will be pressure to retain it in the inventory for much of that period.

Current defence decisions can thus have implications for a nation's defence posture and capabilities some three decades into the future. Defence decision-making is therefore about long-term planning and about planning in a complex, uncertain environment. Like others in the public sector defence decisions are inevitably political choices, dependent upon the policy-maker's judgement. That judgement will be a function not only of experience, general knowledge and intuition, but also of the quantity, quality and organisation of available infor-

mation about the immediate decision problem. While not much can be done about the first three variables, the analytical approach to decision-making holds that a great deal can be done about the information base. It suggests that if the decision-maker can explicitly define his objectives; if he is aware of the available alternatives; if he can recognise the implications of these alternatives; and if he can somehow compare and rate these implications, then better decisions will result.

Cost-effectiveness analysis is one analytical technique that has been widely applied in defence decision-making. In fact, it emerged as a distinctive analytical method in response to defence needs. Its central philosophy is that defence decisions are economic decisions in that they involve the use of scarce resources that have valuable alternative uses. As a result, decisions about force structures, weapons systems and sub-systems should be made on the basis of economic rationality. The following sections of this article consider the characteristics of cost-effectiveness analysis as a distinct analytical approach, its rise as an input to the defence decision-making process, some conceptual and practical problems; a final section offers an assessment of its contribution and future.

THE ANALYTICAL APPROACH

A semantic jungle surrounds discussions of the analytical approach to decision-making. The inclusive term for the general approach has been variously given as policy analysis, systems analysis and operational research. Specific variations have been known as cost-effectiveness analysis, systems engineering, and operational analysis. In some contexts these terms have been used interchangeably; in others they denote some subtle distinction. Typically the usage depends on the analyst. While several authors have attempted to clarify this terminology, the confusion lingers on [1, pp. 2–12; 14, pp. 1–5; 12]. For our purposes it will be recognised that all analytical approaches involve common procedures: careful specification of the objective; enumeration of alternatives; modelling; emphasis on quantitative techniques; and explicit treatment of risk and uncertainty. However, to place cost-effectiveness analysis in perspective and to aid in highlighting its scope, problems and limitations, a brief comparison of this approach with the two others known generally as cost–benefit analysis and operational analysis is useful.

Cost–benefit analysis is an approach to evaluating public sector investments or other activities where both the inputs and outputs can be assessed, either directly or indirectly, in terms of market-established money values. It is easiest to think of this approach in terms of the analogy of the competitive private sector firm which has the single objective of maximising profits. To achieve this objective the firm will undertake investments and arrange its activities such that the difference between the discounted present values of its total revenues and total costs is a maximum. In its calculus the firm will include only those revenues and costs *which actually accrue to and are borne by it.*

Cost–benefit calculations in the public sector differ in terms of the benefits and costs included. Public decision-makers are concerned with *total (or social) benefits and costs.* The criteria for choosing public projects and activities will be analogous to that of the private firm: select those which maximise the 'social profits', the difference between the present values of money-evaluated social benefits and money-evaluated social costs.

Operational analysis is an aid to decision-making which has been widely applied in defence, in other areas of the public sector, and in private industry; it is generally acknowledged to be the approach that is most widely used and the most useful [15, p. 22]. It is one which, at least in its original meaning, deals with a different set of problems from cost–benefit analysis: problems in which neither inputs nor outputs are expressed in money values. Inputs are usually defined in terms of constraints on *specific* resource inputs—manpower, capacity, raw materials, etc. Similarly, outputs are measured in physical terms such as production, waiting time, and inventory levels. The criteria of choice in these problems is typically to maximise or minimise output subject to the set of input constraints. To solve such problems, operational analysts have developed an impressive array of techniques, including mathematical programming, queuing models, and network analysis; indeed the entire operational analysis approach is charac-

Vol. 7
1979

terised by heavy emphasis on mathematically sophisticated models.

Cost-effectiveness analysis includes elements of both cost–benefit analysis and operational analysis. Like the former, inputs are defined in terms of money or *general* resources; like the latter, outputs are measured (where they can be measured at all) in terms of physical or other non-monetary units.

Cost-effectiveness analysis deals with decision problems where outputs cannot be evaluated in money terms. Several areas of government activity approximate this condition—including law and order, justice, some aspects of education and public health—but defence is the classic case.

Because defence outputs are not marketable, costs and benefits are incommensurable and hence the criterion of maximising their difference is not applicable. The decision-making analogy here is therefore not with the profit maximising firm; rather it is with the utility-maximising consumer. Confronted with a money or general resource constraint in the form of his income, the consumer is assumed to allocate his expenditures so as to maximise some concept of well-being or utility, where utility is defined as a function of the quantities of goods and services consumed. An optimum allocation is achieved when the rate at which the consumer *can* substitute between commodities is just equal to the rate at which he *is willing* to do so. Market prices determine the former rate but the latter is subjectively determined in the light of the consumer's utility function. Similarly in defence choices, market-established trade-offs among defence inputs—manpower, equipment, and supplies—must be compared with the substitution possibilities offered by some concept of a defence utility (or defence effectiveness) function. If this is possible, then the criterion for making defence choices is to select that combination of inputs which, for a given budget, maximises effectiveness. An alternative criterion, and one which is logically equivalent, is to select that combination which minimises the cost of attaining a desired level of effectiveness.

The essential difference between cost-effectiveness analysis and operational analysis is the difference between their implicit time horizons. In the former, sufficient time is available for all inputs to be considered variable and substitutable; hence the appropriate constraint is a money or general resource constraint. In operational analysis problems, inputs are in the form of specific resources because insufficient time is available to permit substitution. Thus cost-effectiveness analysis is appropriate to longer-term resource allocation problems; operational analysis is appropriate to short-run problems concerning the efficient use of given resources. But these distinctions should not be overdrawn. Any real world problem is likely to include both general and specific constraints. Whether it is then a cost-effectiveness problem or an operational analysis problem is arbitrary and unimportant.

THE RISE OF COST-EFFECTIVENESS ANALYSIS

The formal application of cost-effectiveness analysis to defence expenditures is a phenomenon of the post-World War II era and has been an accepted management approach within defence establishments for less than two decades. But it is possible to argue that cost-effectiveness of one sort or another always been done. Since he picked up his first club, man has always been fascinated by the effectiveness of alternative weapons and alternative ways of employing them: some would argue that it has been his favourite pastime. Equally, because resource constraints are an inescapable and enduring fact of life, some notion of costs has always shaped his particular choices. Deliberately and explicitly comparing costs and effectiveness, however, emerged only in recent times, both out of the compelling necessity of strategic and technological change and out of the intellectual capability for doing so.

The intellectual lineage of cost-effectiveness analysis can be traced through the independent developments of cost-benefit analysis and operational analysis. Cost–benefit analysis has the longer (which is not to say the more legitimate) history and that history itself can be traced back through both economics and engineering. On the economics side, its theoretical roots are to be found in nineteenth century developments in welfare economics and their later refinement. On the engineering side, England [3, p. 2] suggests that cost–benefit analysis has antecedents in nineteenth century studies of railway location and the later emergence of

engineering economics as a field of study. However, as Dasgupta and Pearce [2, p. 13] point out, it was not until the late 1950s that a reasonably integrated body of knowledge recognised as cost–benefit analysis could be identified.

Operational analysis originated during the early years of World War II when Allied scientists sought means of adapting tactical procedures to technological innovations in military equipment, particularly the efficient use of radar in air defence. Their successes in this area (and elsewhere) led to the creation of permanent operational research groups. Strictly speaking, these were not concerned with questions of efficient resource allocation but with maximisation of the operational efficiency of forces-in-being. In the circumstances, this was appropriate: the ultimate aims of the war effort were graphically clear, and with the economy mobilised for war, the important resource constraints were specific rather than general. But the shift from total war to the quasi-peace of the Cold War era changed the nature of defence choice problems. The planning horizon lengthened. Decision problems were related to force structures and equipment rather than to immediate operational needs. To complicate the situation, the last years of the war had seen a quantum jump in military technology, especially in the destructive power of weapons and in the means of their delivery. Moreover, the objectives in maintaining military forces became less clear-cut than they had been in wartime. Concurrently, as civilian claims on the economy were released from wartime restrictions, defence planners were forced to make choices within severely contracted resource constraints. The conjuncture of these several developments made it clear that new ways of thinking about defence problems were required.

The military organisation most directly affected by these changed circumstances was the United States Air Force; and it was from its associated research organisation, the RAND Corporation, that these new ways of thinking emerged. RAND analysts brought together the ideas of cost–benefit analysis and operational analysis to produce what can now be recognised as cost-effectiveness analysis. This development can be traced to one of their earliest studies, an analysis of offensive bomber requirements, which was done in 1949 [11, p. 2]. In re-examining the conventional tests for selecting military aircraft—traditionally concerned primarily with performance—the RAND analysts sought to apply economic cost concepts. They found, however, that defence accounting data were not structured for such purposes. In particular they found it impossible to obtain directly any measures of the total cost of a weapons system; nor was there available any definition of just what comprised a total system. Cost data were available only in terms of appropriations categories—personnel, real estate, materials, etc.—and then only on a historical basis. Eventually the RAND analysts formulated measures of total aircraft cost, defining in the process the concept of a total weapons system. Among other things, they made the important distinction between one-time and recurring costs of weapons systems and devised a time-phased structuring in terms of development, investment, and operating cost categories. The effect was to shift attention to the full resource implications of selecting a weapons system and away from partial measures such as initial development costs, unit procurement costs, or next year's budgetary requirement.

This led naturally to the idea of total mission or total programme costs, where a programme was identified as a grouping of weapons systems having a common mission goal—strategic retaliatory forces, general purpose land forces, anti-submarine forces, etc.—regardless of the individual service responsible for them. The notion was incorporated in RAND's proposal for the introduction of programme budgeting in the U.S. Department of Defense to facilitate the more rational allocation of defence resources by costing budgetary information in categories related to the actual outputs of the defence establishment and by portraying the time profile of their costs. In short, programme budgeting would provide the information inputs required for cost-effectiveness analysis.

A milestone in the intellectual development of cost-effectiveness analysis was the publication in 1960 of Charles Hitch and Roland McKean's *The Economics of Defense in the Nuclear Age* [7]. Although now dated, the book remains the foundation of defence management thinking. Within a year of its publication, Hitch was appointed Comptroller under

Vol. 7
1979

Defense Secretary Robert McNamara and set about installing both a programme budgeting system and an analytical capability to support it. Since that time, programme budgeting and cost-effectiveness analysis have been adopted throughout the world. Among the early converts were Britain and Canada, both of whose defence ministries were dominated in the 1960s by dynamic and reform-minded ministers: Denis Healey in Britain, Paul Hellyer in Canada.

In retrospect, the success of the McNamara–Hitch innovations is seen to be somewhat mixed. They have fundamentally altered at least the form of defence decision-making and improved the information base upon which defence decisions are taken. But in the late 1960s the analytical approach to defence management acquired a somewhat tarnished image. One reason was the method of implementation. The analytical capability was imposed on the D.O.D.'s military–civilian bureaucracy in the form of wholesale appointments of young, academically-orientated, quantitatively-biased analysts. These so-called 'Whiz Kids' were regarded, particularly by the military, as having an excessive influence on defence policy-making, their calculations frequently over-riding evaluations based on military experience. A second reason was that use of analysis tended to centralise decision-making in the office of the Secretary and consequently diminished the roles of the Joint Chiefs of Staff and the service departments. Thirdly, the analytical approach appeared to depend for its successful application on the personality, abilities and particular working habits of McNamara himself. The systematic approach was a personal monument to his tenure as Secretary: later incumbents would build their own monuments through different management styles. (For a comprehensive review of the changing fortunes of systems analysis in the US Department of Defence see [16].) Finally, the intensification of the Viet Nam War and subsequent American frustration seemed to reveal the impotency of the analytic approach in an actual conflict situation.

Despite these difficulties (and they must be regarded as more than birth pains), cost-effectiveness analysis as an aid to defence decision-making has survived not only in the United States but elsewhere. (An overview of the use of systematic analysis in the UK Ministry of Defence is given in [5].) Its initial penetration, its role as an agent of traumatic change, reaction against it, and finally its absorption into the routine of defence decision-making [16, ch. 10]. Its prestige may have diminished since the mid-1960s, but its acceptability (or toleration) has become much more general. Indeed, it has become a fixture of management in defence and it is seen to be done. Perhaps the most telling stamp of approval is the apparent interest of the Soviet Union in applying the cost-effectiveness approach to its own defence problems [8, pp. 45–46; 10].

CONCEPTUAL AND PRACTICAL PROBLEMS

One of the significant contributions that cost-effectiveness analysis has made to thinking about defence problems is clarification of the meaning and role of cost concepts in decision-making. (The most comprehensive treatment of defence cost concepts is given in [4].) It is a widespread misapprehension that costs, because they can be gauged by the measuring rod of money, are more susceptible to measurement than effectiveness. But a moment's consideration reveals that in principle, both are equally difficult of measurement. From a decision-making point of view, the purpose of measuring costs is to evaluate what cannot otherwise be done if a particular course of action is selected. Costs are therefore measures of the effectiveness forgone; costs and effectiveness are opposite sides of the same coin. Rigorously interpreted, the measurement of costs of a particular option requires that the alternative uses of the resources devoted to that option be enumerated and that the alternatives themselves be evaluated. Costs have no meaning outside these alternatives.

For decisions about allocating resources *among* defence activities finding reliable measures of opportunity costs is problematical, especially where there are constraints on specific resources. Some inputs such as manpower are fixed in supply to military services even though they could be purchased from the market sector at prevailing prices. The expansion of one defence activity at the expense of another therefore requires not only the transfer of general resources in the form of budgetary

appropriations, but specific resources as well. In doing his sums the decision-maker will price these specific resources at their acquisition cost, the cost which represents their opportunity cost to the economy, but not their actual opportunity costs in alternative military employments. The true costs of manpower, for example, may be understated to an expanding activity if manpower has a higher productivity in the declining activities. The result will be a misallocation of the scarce manpower resource among various defence activities, a result which casual observation suggests is all too common. The failure to use a pricing system which reflects the opportunity cost of manpower within defence may lead to its overemployment in activities in which it has relatively low productivity. A similar misallocation of defence resources may occur where the price of an existing facility is considered as zero when its use involves no monetary cost but where the facility has alternative military uses and hence a positive opportunity cost.

A not unrelated problem concerns the application of discounting procedures to defence costs. As suggested previously, time is an important consideration in comparing alternative defence systems. In the absence of discounting, alternative systems are compared on the basis of total service-life costs. But the alternatives are likely to differ in their cost profiles, even if total costs are similar: some may require high initial outlays for development and procurement and low down-stream operating costs while others require relatively low initial outlays and high operating costs. The comparison of differential time-streams of costs requires discounting to present values. As present values are sensitive to the actual rate of discount chosen, the essential problem is to ascertain the appropriate rate. The theoretically appropriate rate is that which reflects both the decision-makers time-preference and his alternative 'investment' opportunities. In defence, decision-makers' time-preference and his alternative 'investment' opportunities. In defence, preferable to forces available in the future. On the other hand, defence decision-makers do have 'investment' opportunities—higher current expenditures on training, maintenance, and development may yield subsequent savings in operating costs. However, the absence of internal capital markets in defence makes the determination of an appropriate discount rate exceedingly difficult if not impossible. The usual procedure is to apply a rate which reflects government borrowing costs, but this may not reflect the time trade-offs available to defence. A preferable approach that has been suggested [4, p. 231] is to compare the total undiscounted costs of alternatives and then test the ranking for sensitivity to a range of discount rates.

Devising sound concepts and measures of cost is obviously a critical element in the design of cost-effectiveness studies. However, the unique feature of such analysis is the need to construct meaningful criteria of effectiveness. The principal difficulties in defining and measuring effectiveness result from the vertical and horizontal proliferation of effectiveness criteria. The reason for this proliferation becomes apparent when one opens the 'black box' of defence to find a hierarchical system of production processes. At the lowest level, where defence has its initial interface with the economy, inputs are acquired from factor markets and combined to produce what might be termed operational outputs. These outputs become the inputs to a subsequent production process whose outputs are the inputs to a still higher process, until some ultimate output is produced.

Consider, for example, the decision to acquire a new maritime patrol aircraft. This aircraft will be combined with other inputs such as manpower, fuel, land facilities, electronic sensing devices and so on to produce the operational output of maritime surveillance sorties. To the commander of the maritime air forces who will be provided with these inputs, the problem (an operational analysis problem) is to maximise, perhaps, the number of sorties subject to the specific resources at his disposal. But this criterion is not the appropriate one upon which to base the selection of new aircraft. Surveillance sorties are but one input into a number of other processes which would include the production of an anti-submarine warfare capability, the production of an anti-shipping capability, the production of protection for fisheries or other off-shore activity, or simply the production of 'presence'. For each of these products there will be a number of criteria of effectiveness and not all of them need be positively related to the number of sor-

Vol. 7
1979

ties, the criterion at the lower level. In the case of anti-submarine warfare capability the criteria would include, among others, speed, range, endurance, detection capability and armament. Some of these criteria would be used to define output of some of the other joint-products of surveillance flights, but their respective contribution to effectiveness in those activities is likely to differ. The relative importance of each criterion will therefore depend upon a higher-level valuation of these several outputs. To complicate matters further, there may exist no single, unambiguous measure for some criteria. Trade-offs among them have to be made, but that combination which, for example, maximises anti-submarine capability at a given expenditure level will be a matter of considerable uncertainty. Modelling of anti-submarine warfare situations using aircraft of various configurations and in combination with alternative mixes of other inputs may reduce this uncertainty by revealing some rough idea of the trade-offs available. But the modelling process is really an attempt to aggregate lower level criteria into higher level criteria, such as submarines detected (or destroyed). If *no* single criterion can be discovered, then the trade-off problem remains.

Even then, an anti-submarine capability is still not an end in itself. At the next higher level in the means-ends hierarchy it is an input for the production of more general defence outputs such as deterrence and damage-limitation. Criteria at this level are even more difficult to define.

In a well-behaved market economy the problem of co-ordinating the values of primary inputs, intermediate outputs and national welfare is supposed to be looked after by the price system. No such automatic co-ordinating system operates in defence. As a consequence, the maintenance of consistency in selecting criteria of effectiveness at various levels of decision-making may be very troublesome. Consider the comparatively low-level problem of selecting a navigation system to be fitted in aircraft assigned to a strike role. Should the criterion of effectiveness be the navigational accuracy of each aircraft such that it may reach its target with some precision? On this criterion the most effective system may, because of its weight or complexity, compromise the range, speed or payload of the aircraft and thereby reduces its strike effectiveness. If so, then the strike effectiveness of the aircraft might be a preferable criterion. But selection of the navigation system on this criterion may reduce the strike effectiveness of the total force because cost considerations limit the number of aircraft which can be fitted. As one ascends the means-ends hierarchy in search of effectiveness criteria, the number of alternative ways of accomplishing whatever the strike role accomplishes will expand. Eventually, the analyst who began with a fairly simple problem may find himself questioning the whole purpose of defence expenditures or indeed the contribution which national security expenditures make to national welfare.

Criteria of effectiveness can expand not only vertically but horizontally as well. In the economist's jargon, defence expenditures have spillover effects. Acquiring a weapons system may have an impact on domestic employment, regional economic activity, research effort, technology transfer and the balance of payments. All or some of these impacts may be important government policy goals and as such potential criteria in the selection of weapons. Indeed, for some countries and under certain circumstances, non-defence criteria may be more heavily weighted by policy-makers than military criteria.

The search for consistency in devising criteria of effectiveness can obviously lead to expansion of even the lowest-level decision problem to the point where it becomes humanly unmanageable. What obviously happens in the real world is that the scope of decisions is constrained. Higher-level decisions are taken as given and the analyst's task becomes one of sub-optimisation. Occasionally, however, he must look up and sideways from his immediate problem.

Given the multi-dimensionality of effectiveness criteria, it is appropriate to think in terms not of individual criteria but rather of military utility or military worth functions which include individual criteria as arguments. This concept underscores the importance of recognising trade-off opportunities among criteria, but it also imparts a false sense of feasibility. Interesting work has been done in utility analysis, but no-one has yet devised a practical utility function for defence (or any other public activity). In the first instance, it is normally im-

possible to achieve direct quantitative measures of effectiveness. How, for example, does one measure the deterrence value of a particular weapon? In practice, analysts search for proxy measures: destructive capability, number of weapons systems, deliverable explosive weight, penetration probabilities, and so on. All of these variables can be quantified, at least approximately, but no single variable can be said to measure the single criterion of defence. In fact, the temptation to use a single measurable variable as *the* criterion of effectiveness is perhaps the most dangerous aspect of cost-effectiveness analysis. The use of such a variable allows the analyst's model to churn out an 'optimal' solution; but the problem solved has little to do with that of the real world decision-maker. So-called 'hard' cost-effectiveness studies are possible only in the very rare circumstances where a quantifiable, uni-dimensional criterion of effectiveness is available.

In a majority of cases therefore the analyst must resign himself to a role similar to that of the consumer researcher: he can cost alternatives and establish criteria of effectiveness, measuring these where possible; but the actual trade-offs are necessarily subjective and hence the final preserve of the decision-maker.

CONCLUSIONS

The conceptual and practical problems of cost-effectiveness analysis pale by comparison with the challenge it faces in simply leading a viable and meaningful existence. The environment of government can be harsh—even hostile—to the idea of economic rationality, an attribute which is neither necessary nor sufficient for government decision-making. Cost-effectiveness analysis is bound to collide with the system of rewards and punishments which actually guides the behaviour of politicians and bureaucrats. This system provides few incentives to allocate energy and time to searching out alternatives and comparing their costs and effectiveness. It may be more rewarding in terms of status and power to seek increased effectiveness by allocating energy and time to the pursuit of greater funding.

Cost-effectiveness analysis aspires to clarifying problems of choice. It must do so on the understanding that all resource allocation decisions in the public sector are ultimately political choices, and that such choices emerge from the rough pulling and shoving of politics and bureaucracy, not from elegantly contrived models of rationality. One virtue of analysis is that it can illuminate the economic costs of political choices, but if analysis is not consistent with the reality of politics then it may not be done at all. A more subtle danger is that cost-effectiveness may be seen to be done, considerable effort may be devoted to it, but in the actual decision-making process it may be ignored, or trotted out when convenient to add an aura of rationality to the support of particular interests.

The principal problem in establishing the relevancy of cost-effectiveness analysis in defence is that until comparatively recently government was both its sole purveyor and its sole client. A broadening of the analysis market would provide a better test of its relevance and would indeed establish whether cost-effectiveness is in fact cost-effective. On the supply side this requires the development of analytical capabilities in extra-governmental institutions. With the availability of alternative analyses of defence alternatives, legislative and public debate on defence issues can take place within the same framework and terminology employed by government. Competition of this sort raises the value of good analysis and at the same time provides incentives for its application. Recent years have in fact seen a flourishing of non-governmental analysis of defence issues. The Congressional Budget Office, for example, has produced several studies focusing on costed alternatives for United States defence policy. (For a discussion of the case for an independent analytical capability for the US Congress see [16, ch. 7]. An example of the CBO analytical studies is given in [17].) Private organisations such as the Brookings Institution and the National Strategy Information Centre have published similar studies. (For example, see [18, 19].) In the United Kingdom, the costed-option approach to the analysis of defence issues has been largely confined to university-centred research institutes such as the Centre for Defence Studies at the University of Aberdeen [6].

Successfully marketing cost-effectiveness analysis requires that careful attention is paid to both quality and costs and that some

Vol. 7
1979

balance is struck between them. Qualitative improvements will be found in the resolution of the conceptual and practical problems outlined in the previous section; particularly significant gains could be expected from improved measures of internal opportunity costs in defence. With respect to the costs of producing analysis an important point is that cost-effectiveness analysis is subject to diminishing returns. No analysis can ever be complete and extending the effort beyond the point where it can no longer be justified is in itself a misallocation of resources. A capability to perform rapid costings of a few realistic alternatives and to subject these to 'quick and dirty' effectiveness comparisons may yield higher returns than more sophisticated and expensive analysis. But perhaps the highest returns of all are obtained from simply thinking of defence problems in a cost-effectiveness framework.

ACKNOWLEDGEMENT

The author would like to thank David Greenwood, Director of the Centre for Defence Studies, Aberdeen for his helpful comments and advice.

REFERENCES

1. BOBROW DB (1969) *Weapons System Decisions.* Praeger, New York, USA.
2. DASGUPTA AK & PEARCE DW (1972) *Cost–Benefit Analysis: Theory and Practice.* Macmillan, London, UK.
3. ENGLISH JM (1968) *Cost-Effectiveness.* John Wiley, New York, USA.
4. FISHER GH (1971) *Cost Considerations in Systems Analysis.* Elsevier, New York, USA.
5. GIBSON JW (1976) Systems analysis in British defence policy making: Some achievements and limitations. In *The Management of Defence* (Ed. MARTIN L). Macmillan, London, UK.
6. GREENWOOD D (1976) Defence programme options to 1980–81. *Aberdeen Stud. in Def. Econ.* **6**.
7. HITCH CJ & MCKEAN RN (1960) *The Economics of Defense in the Nuclear Age.* Harvard University Press, Cambridge, Massachusetts, USA.
8. HITCH CJ (1965) *Decision-Making for Defense.* University of California Press, Los Angeles, USA.
9. HOEBER FP & SCHNEIDER W (1977) *Arms, Men and Military Budgets.* Crane, Ruseak, New York, USA.
10. HOLLOWAY DH (1971) Technology, management and the Soviet military establishment. *Adelphi Pap.* **96**.
11. MASSEY HG, NOVICK D & PETERSON RE (1962) *Cost Measurement: Tools and Methodology for Cost-Effectiveness.* Rand, Santa Monica, USA.
12. NISKANEN WA (1968) Measures of effectiveness. In *Cost-Effectiveness Analysis* (Ed. GOLDMAN TA) pp. 17–32. Praeger, New York, USA.
13. QUADE ES (1964) *Analysis for Military Decisions.* Rand McNally, New York, USA.
14. QUADE ES & BOUCHER WI (1968) *Systems Analysis and Policy Planning: Applications in Defense.* Elsevier, New York, USA.
15. QUADE ES (1975) *Analysis for Public Decisions.* Elsevier, New York, USA.
16. SANDERS R (1973) *The Politics of Defence Analysis.* Dunellen, New York, USA.
17. US Congressional Budget Office (1978) *US Air and Ground Conventional Forces for NATO: Air Defense Issues.* US Government Printing Office, Washington, USA.
18. WHITE WD (1974) *US Tactical Air Power: Missions, Forces, and Costs.* Brookings Institution, Washington, USA.

ADDRESS FOR CORRESPONDENCE: *J Treddenick, Esq. Centre for Defence Studies, University of Aberdeen, Edward Wright Building, Dunbar Street, Old Aberdeen, AB9 2TY, Scotland, UK.*

BLANK
PAGE

Vol. 7
1979

OMEGA The Int. Jl of Mgmt Sci., Vol. 7, No. 5, pp. 469 to 479
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1001-0469$02.00/0

Public Expenditure and Public Accountability

DAVID HEALD

University of Glasgow, Scotland

(*Received May* 1979)

Within a parliamentary democracy, both the Executive and Legislature have key roles in ensuring public accountability for budgetary decisions. The internal budgetary processes of the United Kingdom Government have developed the PESC system of planning public expenditure over a rolling five year period at constant prices. At the same time, the effectiveness of parliamentary financial procedures, far removed from the decision processes, has further declined. The key weaknesses are identified as: (1) the almost complete divorce between parliamentary discussion of expenditure and that of taxation; (2) the manner in which the Government's survival might be imperilled by defeat even on minor items, which limits the activities of Government backbenchers; (3) the unequal knowledge and resources of Government and Parliament, accentuated by British traditions of official secrecy; and (4) the failure of Parliament to seize those opportunities which are available (e.g. debates on the annual Public Expenditure White Paper). Proposals have been made for reform of financial procedures and for a new system of select committees by the Expenditure and Procedure Committees. These are evaluated and given a cautious approval. The emphasis is on the aim of integrating parliamentary debate of the expenditure and revenue sides of the public budget. A single budgetary document is proposed. Multi-year tax forecasts, and the underlying medium term economic assessment, should be published. 'Tax expenditures' should be carefully costed and linked to the relevant expenditure programmes. These reforms would provide a more coherent and comprehensive framework for public debate of budgetary decisions.

THE MEANING OF PUBLIC EXPENDITURE

THE MEANING of 'public expenditure' is less clear than it appears at first sight. Its components include the spending of (a) central government, (b) local government, and (c) part (or all) of the spending of bodies such as nationalised industries, other public corporations and fringe public sector bodies. There are acute problems of international comparison because many national statistics (and those collected by international agencies such as the United Nations, International Monetary Fund, EEC Commission, and the Organisation for Economic Co-operation and Development) concentrate mainly on (a) and (b) (often called 'general government') and lack a comprehensive definition of 'public sector'. The potentially misleading nature of such a limited term of reference has been emphasised by Hood [25] who suggests that one motive for the creation of fringe bodies is to limit the number of public employees *counted as* civil servants. The Bowen report [4] identified as many as 252 'fringe bodies' yet Hogwood [24] has argued convincingly that this is a serious understatement.

The United Kingdon traditionally has had a wider definition of public spending than most countries. Until 1977, public expenditure included the total spending of central and local government and the capital spending of public corporations including the nationalised industries (i.e. operating expenses met by revenues were excluded).

During the 1976 financial crisis and in the run-up to the IMF loan of December 1976, there was much comment about the size of the United Kingdom public sector. Mr Roy Jenkins, then the Labour Home Secretory, argued that public expenditure as 60% of GDP (the then figure) was the 'threshold of totalitarianism'. Without discussing the merits of this political assertion, it seems worth making three

points. First, care must be used when interpreting the public expenditure/GDP ratio. The numerator includes public expenditure on transfer payments (e.g. pensions) which are not part of the denominator, GDP. Second, the ratio is highly sensitive to the definition of public expenditure. The 1976 Public Expenditure White Paper [43] states that the public expenditure/GDP ratio had risen to 60% in 1975/76. Yet the 1979 Public Expenditure White Paper [47] gives the ratio of 46½% for the same financial year. This remarkable transformation of the figures was achieved by moving to a less comprehensive definition of public expenditure (but one nearer to international definitions). Instead of nationalised industries' capital expenditure being counted as part of public expenditure, only their net borrowings from the Exchequer now 'count'. All debt interest which will be met out of trading revenues and not out of taxation is also excluded. The ratio could be reduced still further by taking out social security which is usually treated as a separate fund outside central government in other countries. Third, OECD statistics [31, 32] show that the United Kingdom's level of public expenditure and taxation is broadly in line with that of comparable Western European countries.

During the period 1976–1977, the United Kingdom Treasury apparently encouraged the view that public expenditure was excessively high in relation to other countries [43]; the financial markets then took this literally, and there was a subsequent alteration in definitions to reduce the public expenditure/GDP ratio [45].

The Treasury's *Guide to Public Sector Financial Information* [48] is an invaluable reference source which explains the relationship between public expenditure planning figures, public expenditure on a national income accounts basis and international comparisons. It provides a manual as to the use of published United Kingdom material.

There has been a growing awareness that there is frequently a policy choice between direct public expenditure and 'tax expenditures' (allowances against tax). Examples are child benefits versus child tax allowances, investment grants versus capital allowances, and housing subsidies versus mortgage tax relief. Unlike direct public expenditures, 'tax expenditures' have never been explicitly planned and are not 'charged' against Departments or programmes. Willis and Hardwick [50] have calculated that the yield of UK income tax would have been 40% higher in 1973/74 if there had been no tax expenditures in 1973/74. Willis and Hardwick's call for a 'Tax Expenditure Budget' for the United Kingdom has been partially met by the publication in the 1979 Public Expenditure White Paper [47, table 16] which gives the estimated cost of a large number of tax allowances and lists those on which no valuation can yet be put.

PUBLIC ACCOUNTABILITY

Within a parliamentary democracy, both the Executive and the Legislature have key roles in securing that accountability to the electorate is maintained. The respective role of these two institutions will be a central issue in this article. The effectiveness of parliamentary control over expenditure has so declined that much of parliamentary financial procedure is a charade far removed from the underlying realities of public expenditure. These parliamentary procedures have failed to adapt to the vastly changed circumstances of a large and secularly expanding public sector entirely different from their 19th century origins.

We will now identify a number of recurring themes in this article:

(a) there is a complete divorce between parliamentary discussion of expenditure and that of taxation;

(b) the growth of 'party government' over the intervening period has resulted in decisions on public expenditure and taxation becoming issues of confidence in the Government of the day, thereby severely limiting the ability of backbench members of the Government to take their disagreement with various items anything like as far as voting them down;

(c) there is growing tension between the Executive (Government) and the legislature (Parliament), reflecting the frustration of the latter at its own impotence. Unequal knowledge and resources handicap the ability of the Legislature to perform its supposed task of controlling the Executive. Terence Hig-

Vol. 7
1979

gins MP, a former Conservative Financial Secretary to the Treasury [22], has argued that the parliamentary weakness of the 1974–79 Labour Government encouraged backbench members to flex their muscles. The reports of the Select Committee on Expenditure [8] and of the Select Committee on Procedure [9] have proposed radical changes in financial procedures which are discussed in detail later in this article. These would strengthen the ability of select committees to scrutinise the activities of the Executive. Yet there remains a conflict within the group of reformers who appear to have seized the initiative on this issue. Edward Du Cann MP, the Conservative Chairman of the Public Accounts Committee, regards the strengthening of Parliament's role as a way of eliminating the extensive waste (he sees) in the public sector in order to reduce its size. In contrast, John Garrett, a Labour MP who was formerly a member of the Fulton Committee on the Civil Service, wants a stronger role for backbench MPs as a means of securing more effective democratic accountability and better directed (not less) public spending. These two separate issues are frequently, and sometimes deliberately, confused: first, the desirable size of the public sector; second, the role of Parliament in exercising financial control over the Executive. This division might defeat reform if it were simply seen to be giving weapons to opponents of public spending *per se*;

(d) the ability of Parliament to scrutinize expenditure decisions depends crucially upon the willingness of the Executive to provide background information and material. The British tradition of pervasive official secrecy works against open examination of expenditure decisions. For example, the Executive has consistently refused to publish the PAR (Programme Analysis and Review) studies: it even refused to make available to the Expenditure Committee a PAR study on education which had been made available to OECD and referred to in a published report. Parliament is thereby denied access to information about expenditure *options*. Instead of 'open government', leaks assume a considerable importance. *The Guardian* of 28th February 1979 [18] reproduced parts of a memorandum from Sir Douglas Wass, Permanent Secretary of the Treasury, to Sir Peter Carey, Permanent Secretary of the Department of Industry, stating that in the last four months of 1978 the then Labour Government initiated seven job-saving projects which will involve the country in losses of up to £800m. "The startling and disturbing conclusion is that we have been accumulating prospective losses of real resources at a rate faster than the growth of national income." The official Government response was to order an inquiry into the leak. However, Mr Tony Benn MP [3] suggested that the real question ought to be why the public did not see the other 10,000 letters between Permanent Secretaries. There is a conflict between public access to information and the need for commercial confidentiality on some industrial projects. Yet the present level of secrecy over the whole range of Executive responsibilities hampers the functions of Parliament and inhibits the development of well-informed public opinion;

(e) the structure of the public sector has important implications for the structural forms of public accountability. Part of the case for devolution (e.g. the setting up of an elected Scottish Assembly) has always been the ability of the sub-national unit of government to scrutinise devolved public expenditure more effectively [29]. During the passage of the Scotland Act, there was debate about whether the (United Kingdom) Comptroller and Auditor General and Public Accounts Committee would be able to scrutinise devolved expenditure. The Government successfully resisted this, but the Expenditure Committee has argued that their remit should cover all bodies receiving funds voted by Parliament (i.e. including devolved assemblies, local authorities, nationalised industries, public corporations and fringe bodies). The Procedure Committee [9] has revealed that the post-war Labour Government rejected the view of the then Comptroller and Auditor General that he should be empowered to audit the newly national-

ised industries. This debate raises important issues about the structuring of the public sector itself. For example, the decisions to set up an elected Scottish Assembly or an electricity corporation to manage nationalised assets, would have major implications for the forms and methods of public accountability;

(f) public expenditure decisions have major macro-economic implications and they are taken within some macro-economic framework. The Plowden Report [34] view of public expenditure planning was explicitly Keynesian but the growing influence of monetarist ideas, and the IMF loan of 1976, have led to the adoption of *unconditional* M3 (money supply) and PSBR (Public Sector Borrowing Requirement) targets. The Government's refusal to publish the medium term economic assessment and tax forecasts to match the expenditure plans has been severely criticised by the Expenditure Committee [11]. Ward and Neild [49] have criticised the failure to develop a 'full employment' or 'constant employment' budget and have undertaken their own calculations of fiscal stance.

THE PUBLIC EXPENDITURE WHITE PAPER

The annual Public Expenditure White Paper (due for publication in December but frequently not appearing until January or February) is the visible part of the iceberg of public expenditure decision-making. The 1979 volume [47] is a much weightier and more comprehensive document than the first volume of February 1969 [41]. Much of the credit for these improvements is due to the urgings of the General Sub-committee of the Expenditure Committee whose members have developed substantial technical expertise. Despite occasional irritation at the Sub-Committee's voracious appetite for information (e.g. Joel Barnett MP [2], Chief Secretary of the Treasury, in the 1979 Public Expenditure White Paper debate), the Treasury has responded to many of its requests, particularly more technical ones. The crucial area where little progress has been made is the general exclusion of Parliament from the debate on policy alternatives. This extends over the choice between expenditure options (e.g. the refusal to publish PARs), the refusal to publish the medium term economic assessment or multi-year forecasts of tax revenues, and the cloak of Budget secrecy which precludes debate on taxation options until after the Government has irrevocably committed its prestige to particular changes.

The White Paper presents the Government's decisions on the level and composition of public expenditure over a five year period. Owing to the inflexibility of many commitments in the short term, such a perspective is required if the Government wishes either to reduce the total or to alter the balance of programmes. The growth of public expenditure is planned in the context of the expected growth of GDP in the Public Expenditure Survey (PESC) period. The planning process and the White Paper uses two sets of constant price figures: first, 'in volume terms' which represents the quantity of public services supplied, and second, 'in cost terms including the relative price effect' which reflects the secular tendency of public sector costs to rise faster than those for the economy as a whole.

Although the White Paper is now an impressive document, there remain serious limitations to which we have already alluded. It is an Executive document which does not require parliamentary approval (e.g. in 1979 there was simply a 'take note' resolution in the name of the Government). Although the Government was defeated in the vote on the 1977 White Paper, it subsequently survived a vote of confidence and it remained the Government's planning document. Higgins [22, 23] has argued that MPs should be able to vote on individual items and that this should have real significance. The White Paper debates are generally acknowledged to be uninspiring and badly attended [39]. They are frequently dominated by the 'technicians' (especially those from the Expenditure Committee). They only engage the interest of ordinary backbench MPs if there are major cuts in public spending. There has been an almost complete failure to initiate debate on spending priorities. The constituency role of MPs frequently encourages them to argue for specific increases in expenditure whatever their ideological stance on the aggregate size of public spending. In the 1979 debate, Douglas Crawford (SNP MP for Perth

and East Perthshire) [7] taunted his Conservative neighbour (Nicholas Fairbairn QC, Kinross and West Perthshire) for wanting "cuts in general but not in particular", citing his lobbying for a bypass for the A9 past the village of Blackford. Most MPs are unwilling to make the explicit judgements right across the spectrum of programmes which a more vigorous parliamentary role in public spending programmes would call for.

The form of the White Paper encourages MPs and others to view expenditure decisions in isolation from decisions on revenue. The document gives few hints about the implications of the five year rolling expenditure programme for the future level of taxation. The output growth forecasts of successive White Papers have been over-confident, thereby encouraging the planning of expenditure programmes on the basis of increases in national resources which never materialise. Even the 'middle case', possibly as much aspiration as forecast, often turns out too optimistic. The result is that public expenditure grows faster than GDP and then there are sharp, dislocating cuts in programmes.

The world recession of the 1970s and the loss of confidence in macro-forecasting has shortened the time horizon of public expenditure planning. The third to fifth years are now highly tentative and the introduction of cash limits on programmes has reduced the significance of the plans in volume terms. If the actual rate of inflation exceeds the rate built into the cash limits, there will have to be offsetting cuts in volume expenditure unless the cash limits are relaxed. The new emphasis on monetary aggregates (M3 and PSBR) has directed attention back to cash and away from real resource use. Nicholas Ridley MP [38], a monetarist Conservative who is a member of the General Sub-Committee of the Expenditure Committee, has argued that there should be a move right away from 'funny money' (i.e. constant prices) back to cash.

EXECUTIVE PROCESSES

The Public Expenditure Survey (PESC), of which the annual White Paper is the visible part, owes its origins to the Plowden Report [34] of 1961. This set out a comprehensive framework for public expenditure planning on a a multi-year, constant price basis against the backcloth of the expected growth of GDP. The origins of PESC are discussed in Clarke [6] and Goldman [16] who were both participants. It was intended to replace the hitherto piecemeal consideration of public expenditure.

Plowden recognised that the annual supply procedure was a mechanical constitutional procedure which was unsuitable for the management or control of public expenditure and of negligible use to MPs. PESC represented a parallel, managerialist control system. It would enable strategic decisions to be taken about the aggregate size and composition of public expenditure by the Cabinet on the basis of an institutionalised annual cycle with which other routines (e.g. policy reviews) were synchronised.

During the early 1970s there was considerable confidence that the United Kingdom had made notable reforms in its budgetary processes without becoming engulfed in the traps which had doomed the PPBS (Planned Programming Budgetary Systems) of the Johnson era in the United States (e.g. [13]). Though critical, the Heclo and Wildavsky study [21] did not conceal its admiration for the functioning of the 'expenditure community'. Ironically, the PESC system was by then creaking under the strain of rapidly rising inflation and of frustrated expectations of growth of GDP. Godley [15] dramatised the alleged failings of PESC with his allegation that there was a 'missing £4b' in the financial year 1974/75. Heald [19] noted that Supplementary Estimates for 1974/75 amounted to 32.5% of the initial provision in the original Supply Estimates and that their form did not enable informed judgements to be made about their justification or otherwise. Doubts were cast on the reliability of the revaluation factors being used. Public expenditure in volume terms grew by 19.8% in the three years to 1974/75 (7.5% in that year alone). In cost terms, the comparable figures were 28.3% and 11.5%: these are much higher because of a strongly positive relative price effect over this period.

The Government's response to this growth and to the financial crisis of 1976 was to cut public expenditure and to introduce cash limits [44]. The computerised Financial Information System, already being developed because of

earlier problems with shortfall, was given fresh impetus [5]. This produced much quicker management information and enabled the Treasury to monitor the profile of Departmental spending. The uncertainty and new structure of rewards and penalties generated by cash limits quickly transformed the problem from one of overspending back to shortfall, this time much larger than ever before. The cash limit system was not seriously stretched in the financial years 1975/76 to 1978/79 as the major component of cost (wages and salaries) were easily forecast when calculating cash limits as the non-statutory incomes policy was enforced fairly stringently in the public sector. The breachings of the Government's 5% pay ceiling for the 1979/80 wage round will test the cash limit system. Unless the cash limits are revised upwards, large cuts in volume expenditure will be required. The Chief Secretary's statement of 23rd February 1979 [1] left unclear by how much the Government would revise cash limits upwards as a result of public sector settlements it had approved well above the 5% norm. Cash limits have been used as a substitute for pay policy in the public sector and any upward revision would damage their future credibility. An extremely important development has been the growing use of the Contingency Reserve as a 'kitty' out of which policy changes undertaken during the year have to be paid. Its size is growing, and its status, as money not allocated to programmes, provides a new flexibility at the margin in adjusting policy priorities and in responding to events without exceeding the planned spending total.

Shapiro [40] contended that a fundamental weakness within the PESC system was the emphasis upon 'collegiality in decision-making' whereby groups of Ministers review the various spending programmes. He contrasted this with a system of bilateral discussions between the spending Department responsible for a programme and the Treasury. The presumption of collegiality is that Ministers would forsake their Departmental roles and critically appraise the programmes of other Ministers. Instead of this, alliances have been formed by spending Ministers against the Treasury. The tradition remains strong that a good Minister is one who successfully advances departmental interests and defends them in Cabinet. Alienating other spending Ministers by criticising their programmes would not assist in this prime task. Indeed, Margaret Thatcher MP and Sir Keith Joseph MP, now associated with policies of major cuts in public expenditure, built their political reputations in the 1970–74 Heath Government as resourceful promoters of education, and health and social security spending.

The 'appraisal' function has been much less thoroughly developed than the expenditure planning function. The 1970 Conservative Government displayed initial enthusiasm for appraisal studies as part of its policy of applying business techniques to governmental activities. PAR (Programme Analysis and Review) was launched publicly in 1971 as a generic name for the application of analytical techniques to existing spending programmes and policy options. Successive Governments have refused to publish PARs or even to identify the subjects under study. Occasionally, work undertaken as a PAR has been published without being identified as such (e.g. a cost–benefit study of forestry). It appears that these Conservative reforms had minimal impact and that spending Departments evolved successful counter-strategies (e.g. offering up programmes to which they had no commitment as token sacrifices; suggesting politically sensitive areas; and putting forward for study the programmes for which their justification was most comprehensive). PAR seems to have run out of steam and a Treasury summary [46] of the processes of public expenditure management fails to mention them by name. The 1974–79 Labour Government initiated a series of major policy reviews (e.g. transport, energy and housing) which were not called PARs but which resulted in Green Papers. These documents, however, tended to present considered Government views and not to indicate the full range of relevant policy options.

The secrecy with which such policy studies are conducted excludes both Parliament and public from knowledge of rejected policy options. It makes a balanced assessment of the effectiveness of innovations such as PAR impossible: outsiders lack the requisite information and insiders are not necessarily above professional or institutional biases. Governments restrict access to such information as its release would enable Parliament, the press and public, and especially the Opposition, to question and

Vol. 7
1979

criticise it more effectively. Greater openness would, however, require major changes in public expectations of Ministers: admission of ignorance or indecision would not have to be treated with customary derision or else the Government would find new methods of suppressing information until it had reached its own decisions. Freedom of information legislation, if effective, would provide Parliament with the ability to police Executive decisions on a much more systematic basis. At present, there is no means of following up the Wass memorandum on the seven projects to establish whether Nigel Lawson MP's accusation [28] that they were solely designed to protect Labour marginal seats was justified.

Discontent with overall economic performance in general and with exercise of governmental functions in particular has resulted in frequent institutional change. The Fulton report of 1968 [12] emphasised managerial functions and led to the establishment of the Civil Service Department. The Heath Government set up giant departments (Department of Trade and Industry and Department of Environment which have been partly dismembered subsequently), and the Central Policy Review Staff as a key instrument of central policy evaluation and formation. The Expenditure Committee 1977 report on the Civil Service [8] voiced continued concern at the quality of civil service management. One of the key issues was the relationship between the Civil Service Department and the Treasury. It is indicative of suspicion of the Treasury across the political spectrum that two former Prime Ministers (Heath and Wilson [20, 51]) advocated breaking it up. Heath argued for taking the public expenditure functions from the Treasury and giving them to the Civil Service Department to link these up with responsibilities for personnel and management services. This move would, however, aggravate the lack of integration of spending and taxing decisions which has been a theme of this article. The Expenditure Committee suggested that the management services and public expenditure functions should be reunited in the Treasury. Any fragmentation of the Treasury's functions should be avoided: the task of better control over the Treasury should be achieved through the means of greater openness, less official secrecy and a strengthened role for the select committees.

THE ESTIMATES CYCLE

The estimates cycle constitutes the parliamentary dimension of the system of public accounting and audit. It consists of the Votes on Account (allowing the continuation of public services until the formal estimates are passed); the Supply Estimates and the Consolidated Fund Acts and the Appropriation Act; the Revised and Supplementary Estimates; the Excess Votes; the Appropriation Accounts and the reports of the Comptroller and Auditor General; and the reports of the Public Accounts Committee. It has secured a very high standard of honesty and integrity ('regularity and propriety') in the management of public funds. But it remains a system much better suited to ensure that the King did not spend the money voted for the army on his mistresses than to control a modern government. It produces a mass of detail which either swallows the reader or is left unread. It is slightly more useful to Government Departments than it is to MPs. There is no effective link between the formal financial procedures and PESC which is primarily an administrative process. Cash limits have hitherto been an administrative device separate from the Supply procedure but the 1979/80 Supply Estimates assimilate the cash limits into the Supply procedure and are voted by Parliament. Supply Days have been usurped for general debate on topics initiated by Opposition parties. There are rarely debates on the Estimates themselves. Higgins [23] noted that Opposition backbenchers would be highly unpopular with their own frontbench if they attempted to initiate a debate on the Supply Estimates, thereby depriving the Opposition of an opportunity to attack the Government on an issue of the day. He urged that the Government should make time available for detailed examination of the Supply Estimates on the floor of the House without encroaching upon what has by tradition become regarded as the Opposition's time.

The growing dissatisfaction in Parliament with Westminster's system of public accounting and audit has been clearly voiced by the Expenditure Committee [8, 9] and the Procedure Committee [35]. The role and status of the Comptroller and Auditor General has been singled out as a particular focus of criticism. Normanton [30] supported by Garrett [14].

argued in evidence that the United Kingdom has the lowest status public auditors in Europe. "We are of the opinion that by comparison with other countries our system of public audit is out of date", concluded the Expenditure Committee [8, paragraph 153]. The major criticisms related to the limited range of accounts the Comptroller and Auditor General examines (e.g. which exclude those of local authorities, nationalised industries and fringe bodies); the level and range of qualifications held by staff in the Exchequer and Audit Department and the fact that staffing, recruitment and salaries are controlled by the Civil Service Department, part of the Executive which is being scrutinised; and the narrowness of public audit in the United Kingdon which concentrates heavily on 'regularity and propriety' (i.e. money has been spent on the purposes which Parliament voted it for) rather than 'value for money'. The Expenditure Committee proposed a series of major reforms including the severing of all links between the Comptroller and Auditor General and the Executive so that he and his staff become part of the staff of the House of Commons; the auditing by the Comptroller and Auditor General of all bodies in receipt of public funds voted by Parliament; and his staff servicing the new select committees proposed by the Procedure Committee. After the Government [17] rejected these proposals, the Expenditure Committee produced a strongly worded restatement of its views [9]. There is a close correspondence between the views of the Expenditure Committee and Procedure Committee on these issues, partly through the influence of overlapping personnel.

THE ROLE OF FINANCIAL SELECT COMMITTEES

The Procedure Committee's report was debated in the Commons on 19th and 20th February 1979 and it became clear during the debate that it enjoyed widespread support among backbench MPs and from the Opposition frontbench. The Government were markedly less enthusiastic though Michael Foot MP, Leader of the House, conceded an early opportunity to consider the proposals in detail when he answered the two day debate. The Procedure Committee quickly produced a report [36] suggesting a programme of action and emphasising the setting up of departmental select committees shadowing Departments as the first step.

The most influential of the existing Commons select committees are the financial ones: Public Accounts, Expenditure and Nationalised Industries. The Public Accounts Committee, set up by Gladstone in 1861, is deeply embedded in the financial procedures of the House, and is able to depend upon the resources of the Exchequer and Audit Department. It has impressed high standards of stewardship upon the Executive and its sometimes caustic comments upon lapses carry great weight. It remains, however, a *post hoc* examination of the appropriation accounts and the Public Accounts Committee does not have a role in forward policy making, except for the important one of drawing procedural lessons from its investigations.

The Expenditure Committee was set up in 1971 to replace the earlier Estimates Committee. The intended co-ordinating role of the main Committee has been singularly unsuccessful and the investigations are conducted by the six Sub-Committees. Of these, the General Sub-Committee has established the most accomplished reputation for its work on public expenditure processes. The other Sub-committees have frequently fallen into the temptation of investigating currently topical issues as one-off enquiries without contributing much to the debate on expenditure priorities across programmes. Robinson's study [39] of the early years of the Expenditure Committee reaches deeply pessimistic conclusions about its failure to live up to expectations and about its limited influence on Government.

The Select Committee on Nationalised Industries operates through three sub-committees and reviews the operation, financing and relationships with Government of the nationalised industries and a number of other public corporations. It has tended to see its role as defender of the nationalised industries' independence against the interference of Government departments and it has issued a number of reports severely critical of Government.

The Procedure Committee's recommendations would leave the role of the Public Accounts Committee unchanged (contrary to the wishes of the latter's Chairman, Edward Du Cann MP, who wanted a merger with the

Vol. 7
1979

Expenditure Committee). The Expenditure Committee would be abolished: in a sense, some of its sub-committees would form the nucleus of some of the proposed departmental committees. The Select Committee on Nationalised Industries would be abolished, its functions being taken over by the departmental committees (contrary to the wishes of its Chairman, Russell Kerr MP).

The proposals of the Procedure Committee are generally sound but there are a number of issues and problems which must be faced as a pre-condition to the successful implementation of its select committee proposals. First, there will be additional costs which must be justified with reference to the results of better scrutiny. The select committees will require proper servicing and research resources and there will probably be extensive monitoring of each select committee's activities by the Department it is shadowing. Second, there must be serious doubts about whether existing reports, regardless of their quality, are read outside Government by more than a few Expenditure Committee 'watchers'. Third, Parliament itself does not appear to take select committee reports very seriously. Even those reports which are now debated on the floor of the House (not very many of them) usually attract very poor attendances and the proceedings are dominated by the members of the select committee which prepared the report. Higgins [22], though supporting the development of the select committees' role, warned against the emergence of a new elite of 'technicians'. It is indicative of the status of select committee reports that MPs called for, and the Government appointed, a committee of inquiry (under Mr Justice May) into the prison service [37] just after the Expenditure Committee had produced a comprehensive report [10]. Fourth, there must be a danger that the departmental select committees might be 'taken over' either by the Department being shadowed or by some pressure group interest. This would exacerbate the problem of discussing priorities across departmental and programme boundaries. The ability to take a broader view was the case advanced by Russell Kerr MP [27] against the abolition of the Select Committee on Nationalised Industries.

Supporters of parliamentary reform often either explicitly or implicitly decry the rise of 'party government' and the control that this provides the Executive over the Legislature. They advance a 'consensus' rather than 'adversarial' model of political behaviour. Strengthening the Legislature through the establishment of powerful select committees is seen as a method of reducing the power of the Executive. This view does not appeal to those who view Parliament (and hence Executive power) as an instrument for effecting radical social and economic change. The debate about the Executive's power over the Legislature cannot entirely be separated from the debate over the state's role in economic and social life. The conflict between the reformers who want simply 'less government' and those who want 'better government' may still abort reform in a political climate of bitter controversy about the role of the state and the size of public expenditure.

CONCLUSIONS

The steady process of improving the annual Public Expenditure White Paper should continue. The most crucial change is to move towards the integration of the parliamentary debate of the expenditure and revenue sides of the public budget. This involves the publication of multi-period tax forecasts and the underlying medium term assessment, and the accurate costing of tax expenditures which should be linked to the relevant expenditure programmes. Parliamentary and public attention now focuses upon the melodrama of Budget Day rather than upon the longer term decisions on public expenditure which shape the amount of room for manoeuvre which the Chancellor of the Exchequer possesses in any particular fiscal year. Although there would be problems of timing to be resolved, a single budget document (covering both expenditure and revenue) should replace the existing Public Expenditure White Paper, Financial Statement and Budget Report, and Memorandum of the Chief Secretary to the Treasury on the Supply Estimates. This single document should be backed up by appendix volumes. There is an important role for guides to the system and to published documentation: the Treasury's *Guide to Public Sector Financial Information* [48] is a welcome addition. Unfortunately, however, the guide to the methodology of the Public Expen-

diture White Papers [42] is seriously out of date. Much greater openness (e.g. the publication of PARs) can assist in channelling debate in constructive directions. Establishing the scale of public expenditure and the priorities within it ought to be at the heart of the democratic political process: it ought not to be left entirely to the Executive nor to a group of 'specialist' MPs.

The Expenditure and Procedure Committees have made valuable suggestions about the evolution of the select committee system. These reforms can be successful in reviving parliamentary scrutiny of expenditure decisions provided certain conditions are met. They will fail if they are envisaged as a means of 'depoliticising' public expenditure decisions and of undermining 'party government' in favour of a 'consensus' style of government. They will also fail if the purpose becomes *solely* to discredit government activity and to reduce the size of the public sector. Both these developments would alienate MPs otherwise sympathetic to the goal of securing more effective accountability for public expenditure decisions. Public expenditure, bureaucrats and politicians have been the targets of a mounting series of skilful attacks from the 'radical right' [26]. Political developments such as California's tax-cutting Proposition 13 suggest that the post-war consensus in favour of 'welfare capitalism' has been eroded and that sharp political conflicts over the role of the state are in prospect. This will limit the 'consensus-generating' role of select committees. Nevertheless their investigative and information-gathering roles provide valuable inputs into the party political conflict.

PESC provides the framework for Executive decisions on the size and mix of public expenditure. If opened up to wider scrutiny, Parliament could play a more effective role. Macroeconomic circumstances and policy have a dominating influence over the decisions on broad aggregates. The 1974–79 Labour Government latterly committed itself to M3 and PSBR targets which were not conditional upon the level of economic activity. The present author would support the [33, 49] case for judging fiscal stance by the constant employment budget balance. A critical issue facing PESC is the future of incomes policy and cash limits. Some form of incomes policy is inevitable in the public sector especially in tax-financed public services. Cash limits have been acting as a surrogate for a public sector incomes policy but it is as yet unclear as to whether they will hold intact through a renewed bout of wage inflation. If they do not, they will be discredited. If they do, the PESC system of volume planning will exist only in name but not in substance.

REFERENCES

1. BARNETT J MP (1979) *Hansard*, 23rd February, column **335**.
2. BARNETT J MP (1979) *Hansard*, 19th March, column **1108**.
3. BENN AW MP (1979) *Speech to annual conference of the Scottish Council of the Labour Party*. Perth, 9th March 1979.
4. BOWEN G (1978) *Survey of Fringe Bodies*. Civil Service Department, London, UK.
5. BUTLER FER & ALDRED K (1977) The Financial Information Systems Project. *Management Services in Government* pp 77–87.
6. CLARKE R (1971) *New Trends in Government*. Civil Service College Studies **1**, HMSO, London, UK.
7. CRAWFORD D MP (1979) *Hansard*, 19th March, column **1172**.
8. EXPENDITURE COMMITTEE (1977) *The Civil Service*. Eleventh Report of session 1976/77. HC **535**, HMSO, London, UK.
9. EXPENDITURE COMMITTEE (1978) *Response to the Government's observations on the Committee's report on the Civil Service*. Twelfth Report of Session 1977/78. HC **576**, HMSO, London, UK.
10. EXPENDITURE COMMITTEE (1978) *The Reduction of Pressure on the Prison System*, Fifteenth Report of Session 1977/78, HC **662**, HMSO, London, UK.
11. EXPENDITURE COMMITTEE (1979) *The Government's Expenditure Plans* 1979–80 *to* 1982–83, (Cmnd 7439), Fourth Report of Session 1978–1979, HC **237**, HMSO, London, UK.
12. FULTON Lord (1968) *The Civil Service*. Report of Committee, Cmnd 3638, HMSO, London, UK.
13. GALNOOR I (1974) 'Reforms of public expenditure in Great Britain'. *Can. J Publ. Admin.*, **17**, 289–320.
14. GARRETT J MP (1978) *Public Accountability, the Expenditure Committee and the Public Accounts Committee*. Printed as evidence in Procedure Committee (1978) Vol. **III**, Appendix 44, 141–144, HC **588-III**, HMSO, London UK.
15. GODLEY WAH (1975) *Public Expenditure* 1970–71 *to* 1974–75. Printed as evidence in First Report from the Expenditure committee, Session 1975–76, The Financing of Public Expenditure Vol. **II**, HC **69-II**, p 224 HMSO, London, UK.
16. GOLDMAN S (1973) *The Developing System of Public Expenditure Management and Control*. Civil Service College Studies **2**, HMSO, London, UK.
17. GOVERNMENT (1978) 'The Civil Service. Government observations on the Eleventh Report from the Expenditure Committee', session 1976–77, HC **535** of Session 1977–78, HMSO, London, UK.
18. THE GUARDIAN (1979) 28th February. Page 1. London, UK.

Vol. 7
1979

19. HEALD DA (1975) *The Control of Public Expenditure.* Printed as evidence in First Report from the Expenditure Committee, Session 1975–76, The Financing of Public Expenditure Vol. II, HC **69-II**, pp 199–203, HMSO, London, UK.
20. HEATH E MP (1978) Oral evidence, p 760. Expenditure Committee, HC **535-II**, HMSO, London, UK.
21. HELCO H & WILDAVSKY A (1974) *The Private Government of Public Money.* Macmillan, London, UK.
22. HIGGINS T MP (1978) Parliamentary control of public expenditure—a back bench revolt? *National Westminster Bank Q. Rev.* 26–35.
23. HIGGINS T MP (1979) *Hansard*, 19th March, column **1189**.
24. HOGWOOD B (1979) *The tartan fringe: quangos and other assorted animals.* Paper given at joint conference of Political Studies Association's Local Politics and UK Politics Work Groups, Glasgow, UK.
25. HOOD CC (1978) 'Keeping the centre small: explanations of agency type', *Polit. Stud.* Vol. **XXVI**, 30–46.
26. Institute of Economic Affairs: Hobart Papers and Occasional Papers, including NISKANEN W (1973) *Bureaucracy: servant or master*; TULLOCK G (1976) *The vote motive*; and HARRIS R & SELDON A (1977) *Rates or prices.*
27. KERR R MP (1979) *Hansard* 19th February, columns **78–87**.
28. LAWSON N MP (1979) *Hansard*, 19th March, column **1225**.
29. MACKINTOSH JP MP (1968) 'The Devolution of Power' Penguin, London, UK.
30. NORMANTON EL (1978) *Public Accountability.* Printed as evidence in Procedure Committee (1978), Vol. **III**, Appendix 43, pp 131–140, HC **588-III**, HMSO, London, UK.
31. OECD (1978) Public Expenditure Trends OECD Studies in Resource Allocation, No. **5**, Paris, France.
32. OECD (1978) *Revenue Statistics of OECD Member Countries*: 1965–1976. Paris, France.
33. PARDOE J MP (1979) *Hansard* 19th March, column **1169**.
34. PLOWDEN Lord (1961) *Control of Public Expenditure.* Comnd **1432**, HMSO, London, UK.
35. PROCEDURE COMMITTEE (1978) *First Report of Session* 1977–78. HC **588**, HMSO, London, UK.
36. PROCEDURE COMMITTEE (1979) *First Special Report of Session* 1978–79. HC **267**, HMSO, London, UK.
37. REES M MP (1978) *Hansard*, 17th November, column **386**.
38. RIDLEY N MP (1979) *Hansard*, 19th March, column **1192**.
39. ROBINSON A (1978) *Parliament and Public Spending* Heinemann, London, UK.
40. SHAPIRO D. (1979) *Seminar at University of Glasgow.* 12th February.
41. TREASURY (1969) Public Expenditure 1968–69 to 1970–71. Cmnd **3936**, HMSO, London, UK.
42. TREASURY (1972) *Public Expenditure White Papers: Handbook on Methodology.* HMSO, London, UK.
43. TREASURY (1976a) *Public Expenditure to* 1979–80. Cmnd **6393**, HMSO, London, UK.
44. TREASURY (1976b) *Cash Limits on Public Expenditure.* Cmnd **6440**, HMSO, London, UK.
45. TREASURY (1977) *The Government's Expenditure Plans.* Cmnd **6721-I**, HMSO, London, UK.
46. TREASURY (1978) The management of public expenditure. *Economic Progress Report* supplement, October. HMSO, London, UK.
47. TREASURY (1979a) *The Government's Expenditure Plans,* 1979–80 *to* 1982–83. Cmnd **7459**, HMSO, London, UK.
48. TREASURY (1979b) *Guide to Public Sector Financial Information.* No. **1**, HMSO, London, UK.
49. WARD TS & NEILD RR (1978) *The Measurement and Reform of Budgetary Policy.* Heinemann for the Institute for Fiscal Studies, London, UK.
50. WILLIS JRM & HARDWICK PJW (1978) *Tax Expenditures in the United Kingdom.* Heinemann for the Institute for Fiscal Studies, London, UK.
51. WILSON JH MP (1978) Oral evidence, p 780. Expenditure Committee (1977) HC **535-II**, HMSO, London, UK.

ADDRESS FOR CORRESPONDENCE: *DA Heald, Esq, Department of Management Studies, University of Glasgow, 25 Bute Gardens, Glasgow G12 8RS Scotland, UK.*

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 481 to 487
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0481$02.00/0

EDITORIAL

Evaluating Management Education

MANAGEMENT education is now big business. I have not seen any recent statistics on expenditure in this field in various countries, but with extensive programmes in operation in thousands of business schools and other educational centres, the total annual figure must be quite phenomenal, not only in the running and capital costs involved, but in the time invested by participants in the numerous courses on offer. It is, therefore, only natural that people should want to investigate the return on this large investment and try to determine whether it manages to achieve what is expected of it.

In this respect management education is no different from education in other disciplines, and an enormous amount of research has been directed towards the evaluation of teaching methodologies, assessment of performance, correlation of achievements on courses with other educational attainments and/or personal attributes, as well as follow-up studies which attempt to trace subsequent career development. Thus, the literature in the field of education research (and methods of investigation in the social sciences for that matter) should provide a most useful background for anyone interested in evaluating management education, and the experience of previous research workers should prove invaluable in determining the kind of investigation to be undertaken.

Now, supposing that you are responsible for running a particular management course, how would you go about assessing its effectiveness and how would you produce an optimal design for the course, so that the resources at your disposal can be utilized in the best possible way? If you constantly preach to your students that management is to do with the use of resources, that the performance of a managerial system always needs to be monitored and evaluated, then surely the activities of your own centre should be amenable to the same kind of treatment.

You start, presumably, by specifying the objectives of the course and the expectations of the students who come on it (and/or their sponsors). You then have to devise a methodology for measuring the major

attributes that delineate these objectives, so that a comparison between aspirations and reality can be undertaken. In the final phase of such an investigation you would consider ways of modifying your course programme with the aim of achieving either better results, or the same results more economically, or both. Needless to say, in this last phase you would also contemplate on the rationale and viability of the stated objectives, which—in the light of your experience—could perhaps be recast and thereby have implications on the future design of the course.

This conventional approach seems straightforward enough, until you try to implement it. It is then that you find innumerable hurdles, each of which can be sufficiently serious to put the whole investigation in jeopardy. I shall mention only some of the problems that are likely to be encountered, but these may suffice as an indication of the enormity and complexity of the task involved.

First, how do you find out what the objectives are for a given course? If you are responsible for running it, presumably you know what you have set out to achieve, and if you are an outside investigator, you can approach the man in charge and ask him to state his objectives. But not unlike other organizational systems, an educational centre rarely responds to such probings with clear-cut answers. Understandably, managers are often reluctant to commit themselves, and even when they genuinely try to cooperate they find difficulty in identifying the objectives of a given system, except in a very narrow functional sense (such as keeping within a prescribed budget or meeting specified targets). Then, there are the customers to consider; they avail themselves of the product or the service on offer because of certain expectations that they have. Should these expectations be explicitly sought and incorporated as part of the objectives-set by which performance is to be judged? It appears that the problem for the investigator is not only to resolve the possible incompatibility between the objectives of management and those of its customers, but to determine what weight to give to each of the sub-sets as part of the performance assessment exercise.

Clearly, any ranking of objectives for this purpose, or assigning weights to them, is open to challenge, and the only sensible course open to the investigator is to maintain a multi-objective set, so that performance can be compared against each objective separately. The implication of such an approach is that it would not then be possible for the researcher to pronounce a verdict as to whether the system's performance is good or bad: it may be good as judged by one criterion but abysmal as seen by another. It would not be possible in many cases even to say whether any contemplated changes, say in a course curriculum, would be an improvement or not, since it would be rare for the system to yield greater attainments against *all* the objectives in question, and more often than not an improvement in one dimension

Vol. 7
1979

would be offset by diminution of performance in another. Similarly, if two courses were to be judged by the same set of criteria, it would not be possible to say which were the better, unless one scored more than the other on every single yardstick. When people do express preferences in a multi-objective situation they invariably do so because they judge some criteria to be more important than others; in other words, they resort to ranking or weight attribution, but then people's preferences vary and a universally accepted judgement is obviously difficult to come by.

The measurement phase is even more intricate. Take, for example, the commonly expressed statement that one of the objectives of a course is to impart substantive knowledge to the student. How can we determine whether such an objective has been achieved? If we mean by substantive knowledge that there is a finite set of facts, figures and formulae which the student needs to know, then his level of learning can be ascertained through the application of one or several rather mechanical tests, and even then it would be necessary to repeat the tests at various time intervals after graduation from the course to determine the level of knowledge retention. But many people will take objection to such a narrow view of the learning process and will argue strongly that interpretation, appreciation, judgement, decision-making and ability to lead and negotiate are qualities that need to be inculcated or nurtured; these attributes are much more difficult to measure by a series of mechanical tests, and when such tests are devised, the results are open to many possible interpretations.

The alternative to objective testing is to solicit views and try to establish what attitudes are held by various interested parties, such as teachers, students, sponsors, employees, bosses, underlings and so on. These attitudes are usually sought by detailed questionnaires backed up by some interviewing. I have discussed the many pitfalls in such a process of inquiry elsewhere [1] and it would not be appropriate to repeat the arguments here; suffice it to say that, even if a true and unbiased picture of people's attitudes can be obtained from such an investigation, it may only be regarded as a rather incomplete profile with severe limits on its value for decision making purposes.

In view of such monumental problems of definition, measurement and interpretation, any attempt to evaluate management education as a whole must be regarded as a rather ambitious and awesome enterprise. It seems more sensible for an investigator to devote his talent and effort to the evaluation of a particular educational programme, in the hope that such a case-study will not only provide useful feedback to those in charge, but will highlight at least some of the methodological difficulties alluded to above, and possibly many others.

This is precisely what was done in a study recently published by RM Hogarth [2]. His report concerns the particular programme of CEDEP (Centre Européen d'Education Permanente at Fontainbleau

in France), described as a half-way house between an internal company course and an external course at a business school. The CEDEP course (or 'cycle') consists of eight residential periods at the Centre, their total duration being 90 days, spread over $2\frac{1}{2}$ years. The students covered by the study were all managers, sponsored by eight comparatively large participating companies (five French, the other three from Belgium, Switzerland and Denmark); the study involved some 730 students (admitted in the 1970s to eleven cycles) with wide ranges of age (from under 30 to 50 and over), experience and managerial seniority, as well as educational background.

From the 1972 CEDEP brochure Hogarth quotes the following four objectives of the programme [2, p. 5]:

"(1) To allow the acquisition, deepening, and development of basic knowledge, concepts, and techniques necessary for the understanding and rational analysis of problems of management and the managing of business.
(2) To develop in the participants decision-making skills appropriate to complex situations involving different levels of risk and uncertainty.
(3) To develop attitudes of perception and adaptation to the socio-economic environment and to create awareness of the importance and role of sociological factors both outside and within the company.
(4) To develop attitudes of support and participation in formulation of company objectives and strategy in order to ensure profitability, development, and competitivity."

Vol. 7
1979

While it is understandable that objectives in such circumstances have to be expressed in these vague terms (indeed, one would be hard put to define them much more explicitly), they inevitably sow the seeds of confusion when the design of a measurement methodology is undertaken. I have already alluded to the problem of measuring the degree of acquisition of knowledge (objective 1), and furthermore views do vary as to what basic knowledge, concepts and techniques are necessary for understanding business and management problems. The second objective may seem amenable to objective testing, but the design of such tests and their application are fraught with numerous difficulties, while the last two objectives involve the development of attitudes, and the use of questionnaires to examine this area becomes a natural choice. In the event, Hogarth and his research associates resorted to questionnaires as their major investigative tool in most of the studies they conducted. They did look at some physical variables as well, such as the amount of paper and course material thrown away, bar receipts, telephone calls, and the frequency of eating out; these were regarded as useful indicators of stress and deviation from average behaviour, but in the main the studies concentrated on attitude sur-

veys: reactions of participants to the programme, their expectations from and perceptions of CEDEP and its role in the sponsoring companies, selection policies for sponsoring participants, perception of what had been learnt, of job-related behavioural changes, and of decision making style, and many others too numerous to list here.

If the portfolio of studies undertaken by Hogarth and his coworkers had to be summarized, it would relate to various effects induced by CEDEP with respect to the following categories [2, p. 223]:

"1. Acquisition of knowledge in different disciplines;
2. Acquisition of a more global approach to problems;
3. Increased self-confidence and job efficiency;
4. Greater understanding of and sensitivity to human relations;
5. Decision-making style and attitudes towards risk and uncertainty; and
6. Attitudes and feelings towards the company."

Clearly, the degree to which conclusions can be drawn from these studies does vary from category to category, and some of the participants' responses were both interesting and illuminating. But my general impression of the results is that they are what you would expect them to be: many satisfied customers, some disappointment, some criticism of the programme and the teachers, an appreciation of the opportunity to learn new subjects, a feeling of increased self-confidence, a claim for better understanding of the behaviour of subordinates compared with that of supervisors, and variations in the culture and environment of the sponsoring companies. The heterogeneous intake to the CEDEP programme probably contributed to the spread of some responses, and the value of the course to the individual participant must be conditioned by his background and experience.

The questions that we have to ask ourselves is whether the mammoth exercise was really worthwhile. I have no doubt that the researchers have learnt a great deal from the experience and have gained a better understanding of the many difficult problems that they encountered. But whether the rest of us have been enriched by their experience is debatable.

Take as an example (and it is only one of many) the study concerning the participants' own assessment of what they felt CEDEP actually promoted most or least (they were asked to choose from a list of 25 possible objectives), and the results are ranked below [2, pp. 206, 224]:

Objectives selected most

"1. Generalist outlook
2. Good communication skills—especially listening
3. Sensitivity towards the social environment
4. Ability to face up to complexity (e.g. of problems, issues etc.)

5. { Ability to tolerate ambiguity and uncertainty
   { Ability to manage change
7. { Skill in interpersonal relations
   { High degree of analytical and conceptual skills"

Objectives selected least

"1. { High technical job expertise
    { Idealism
3. Ambition
4. High motivation and commitment to work
5. Leadership ability
6. { Courage
   { Humility
8. Pragmatism"

With the type of programme mounted by CEDEP these results are not in the least surprising, and if the rank ordering had been somewhat different it is doubtful whether it would have made much difference. There are probably two reasons for this. The first relates to Hogarth's own reference to the proposition that research methodology in the social sciences is judged by its properties of *validity*, *reliability* and *generalizability*. I do not wish to question the validity and reliability of Hogarth's results, both because he has evidently taken as much care as was possible under the circumstances to minimize or to isolate many factors that tend to undermine one's confidence in the design of studies of this kind, and also because the accuracy of the results is far from being crucial to the reader's assessment of the programme. As for generalizability, there can hardly be a claim for that, and irrespective of how detailed and meticulously organized a case study is, one swallow does not make a summer. Furthermore, as Hogarth himself admits "evaluation is, by definition, *subjective*, the criteria, purposes, and value assessments concerning a programme can vary from individual to individual. There is no logical reason why outcomes of an evaluation should be the same for the different parties concerned", particularly since "objectives are not explicit" [2, p. 8].

The second reason for having certain doubts about the value of such detailed research is that being a kind of a "quasi-experiment" it offers an opportunity for model building and for proposing hypotheses. Hogarth cites Campbell's view that "such 'quasi-experimental' research designs can be valid provided that one can show that '*plausible* rival hypotheses' which might also account for observed effects can be disproved" [2, p. 239]. Alas, that is the rub in exercises of this kind, since alternative hypotheses do remain plausible and are often difficult or impossible to refute.

SAMUEL EILON
Chief Editor

Vol. 7
1979

REFERENCES

1. EILON S (1979) *Aspects of Management*. 2nd ed., ch. 8. Pergamon Press, Oxford, UK. (Also in *Omega*, **5**(3), 227–233).
2. HOGARTH RM (1979) *Evaluating Management Education*. John Wiley, Chichester, UK.

1. 7
79

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 489 to 491
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0489$02.00/0

FEEDBACK

A Question of Size

PROFESSOR EILON in a recent Editorial [5] has raised so many important issues that it would need a symposium report to do justice to the subject. I comment here on only a tiny part of the whole, but I believe it to be a fundamental starting point for inquiry. My dual text is 'size and complexity impose strain on the information system' and 'size tends to impede communications'.

In any large organisation, communication both oral and written suffers distortion of many kinds as it passes up and down the chain, due to omissions, errors and, more importantly, accretions (see, for example [6]). This distortion is also affected by the different world-views of individuals. Eden [4] has pointed particularly to the different views of the organisation, and its function, that are held by those who operate within it. Theoretically, smaller organisations suffer from similar difficulties but, if personal interchanges are sufficiently frequent and cooperative, language problems become less obtrusive as a common culture is forged. But it still must not be forgotten that the culture developed within the organisation is, for each individual, only a part of a wider culture. This necessarily affects views of what are satisfactory aims, goals and ideals for guiding action. We are all probably aware that to create family harmony and cooperation, parents and children have to strive for a sharing of, or at least an empathy with, each group's wider culture.

The same difficulties beset information. What is information for one part of an organisation needs to be modified and expressed in a different context and a different language for another part. These parts may be individuals or identifiable, coherent, groups. Ackoff and Emery [3] give a possible starting-point for understanding the inherent complexity of information in a behavioural setting.

Conflict, in the most general sense, is inevitable. It can be, often by cooperative intent, both constructive and beneficial. How it develops, and how it is controlled or resolved, depends on communication and shared information. Whether or not, for any organisation, we come to the conclusion that 'big is best' or 'small is beautiful', we shall be likely to be happier with that conclusion if we know more about the particular problems of communication and information handling. This implies appropriate studies of the existing system and suitable theoretical extrapolation to possible new systems.

I suggest, therefore, that an essential component of any organisation should be one that is concerned with monitoring and inquiring into the ways in which we communicate, into the form of our information systems, and into our understanding of the cultures of the organisation and its separate parts. Such a component should provide basic ideas, not only for organisational development and information handling, but also for personnel management and for industrial and public relations.

All we need are geniuses—or, as I and many others [1, 2, 7] would have it, the generators of a revitalised operational research.

REFERENCES

1. ACKOFF RL (1979) The future of operational research is past. *Jl Opl Res. Soc.* **30**(2), 93–104.
2. ACKOFF RL (1979) Resurrecting the future of operational research. *Jl Opl Res. Soc.* **30**(3), 189–199.
3. ACKOFF RL & EMERY FE (1972) *On Purposeful Systems*. Tavistock, London, UK.
4. EDEN CL (1978) Operational research and organization development. *Hum. Relat.* **31**(8), 657–674.
5. EILON S (1979) A question of size. *Omega* **7**(1), 1–7.
6. LAING GJ (1973) Limits to Growth in Business Organisations. PhD Thesis, London School of Economics and Political Science, London, UK. (For an abbreviated version of his work on communication see LAING GJ (1978) *Reliability and human communication*, 5th Symposium on reliability Technology, September. National Centre of Systems Reliability, Warrington, 203–219).
7. MITROFF II (1979) Zen and the art of implementation: speculations on a holistic theory of management. *Jl Enterprise Mgmt* **1**(2), 55–61.

KC BOWEN

Head of Research
Ministry of Defence
Defence Operational Analysis Establishment
Parvis Road
West Byfleet
Surrey KT14 6LY
UK

The Vroom–Yetton Leadership Model

EILON's editorial makes sense to me. He scholarly points out the positive and negative values of the model. I agree that it is a somewhat stochastic model. I, as a manager, may not have the *time* to draw a decision tree! But the factors or issues that Vroom and Yetton ask us as managers at least to consider in the leadership

process (note that I said 'process') are well thought-out and exceptionally perceptive. I myself find the model to be prescriptive, relatively normative, and somewhat mechanistic. As a manager, all that I must do is to ask myself seven questions to decide how to behave in a leadership role. But, in seven of the outcomes of the decision tree (1, 3, 4, 6(b) 7, 8 and 10), I still have *another* decision to make. And! the answers to these questions are most likely a function of *me* (the manager).

Leadership is a role. A role is a set of expectations about how a person ought to behave. If I am a professor, there are expectations of others about how I ought to behave. If I am a father of three children, there are some societal expectations about how I ought to behave *vis-à-vis* these children. Leadership is no different. It is a role.

There are, for me, four roles in the leadership process. These are:

- the expert—the person who gives information and expertise to his peers, superiors, and subordinates. *Don't* worry about looking around for experts! Everyone is an 'expert'.

- the facilitator—the person who asks for and elicits information and feelings from others. For example, "Jim, I know that you have knowledge about this issue we are talking about and I also notice a scowl on your face. Would you help us with your expertise and would you tell us why there is a scowl on your face?" Here, I am a facilitator. *Worry* about being a facilitator. It is a difficult and sometimes demeaning role.

- the coordinator—the person who brings seemingly divergent points of view and divergent feelings together to make an integrated whole. "Mike—you have proposed plan *A*. Jim—you have proposed Plan *B*. How about taking parts of plan *A* and *B* and drawing up plan *C*?" I'm a coordinator here. *Worry* about being a coordinator. It is a difficult and demanding role.

- the morale builder—the person who knows when to tell a joke to ease the leadership process, who knows when to say, "Hey, we've been at this meeting for three hours! Let's go have a stout." Here, I'm a morale builder. *Don't* worry about seeking out morale builders. They will be there in the leadership process.

Four roles in the leadership process—who is a leader? A leader in my opinion is the person who:

- recognizes that there are these four roles in the process.
- recognizes that *all* four must be played for successful leadership.
- is able to play any of the four at any point in time.
- is able to play more than one role at any point in time.

Does this mean that the 'leader' may change over time in any decision-making group? Yes. The leader is the one who provides the group with the role(s) that is especially lacking at any point in time. The 'leader provides what is needed most (*vis-à-vis* role performance) in the leadership process.

As leaders, we all have preferred leadership *styles*. I am suggesting that, despite preferences in leadership style, a successful leader modifies his leadership *behaviour* to suit or fit the problem at hand, the people involved and the situational variables.

For an earlier, yet still vitally current look at the leadership process, see Tannenbaum and Schmidt [1].

Vol. 7
1979

REFERENCES

1. Tannenbaum & Schmidt (1958) How to choose a leadership pattern. *Harvard Busin. Rev.* March–April, 95–101.

Department of Management & Organizational Behaviour
Babson College
Babson Park (Wellesley)
Massachusetts 02157
USA

John J Morse

Ackoff's Fables

Presumably, *OMEGA* is so-named to represent the spectrum of ideas, Alpha to Omega (Rev. xxii, 13). Since there is no fixed number of ideas in this interval, we wait for our issues of *OMEGA, MANAGEMENT SCIENCE, JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY* etc., expecting further revelations.

Great insights have always been in short supply. Sometimes they arrive in mathematical or algorithmic form, capturing the logic of a process of reasoning about relevant relations. Aesop's Fables provide a different kind of model for communicating a logical analysis. The Harvard cases are longer but less robust than Aesop's pithy stories. Both are story-telling methods which Eilon describes as "a most successful recipe for passing on experience to the uninitiated" [2].

To this literature we now add Ackoff's Fables [1]. Is this latest text yet another volume added to a plethora of clever stories with trick endings. I do not think so. Ackoff has an original way of perceiving situations, an unconventional framework for analyzing and synthesizing systems and a broad set of experiences to draw upon. What he writes about cannot be captured in mathematical or algorithmic form. It would be a loss if he decided to respect rigor and thereby keep his fables for lectures alone.

I am sympathetic to Eilon's wariness of cases and stories. The license one takes in constructing a story is evident for all of us, but *caveat emptor* has always applied to the reader and listener. There is no consumer protection agency that can demand total veracity from a story teller to protect the unwary. And we all learn to judge for ourselves what part is literary, what part rhetoric, what part countable and what part imaginative. That exercise alone may justify the story mode of communicating ideas and concepts. The buyer beware is

always aware, and so I do not worry for the wary. I would rather read about insights which "If they are not true, they ought to be" than about perceptions which are true, but mundane, ordinary, expected and dull.

I compliment Eilon for his head-on analysis of this form of communication. His statement that "objectives are not what they seem to be" and his clarification of constraints apply to Aesop as well as Ackoff. But the existence of concepts for better ways to listen and appraise does not negate the benefits to be derived from stories.

Perhaps story telling is 'alluring' but that which is alluring is so in the eyes of the beholder. Surely an intelligent listener can differentiate between false allure and the allure of truth. Stories do not have to be deficient because they are simple. They may be deficient because they are complex. As Stuart Chase observed, "common sense is that which makes us think the world is flat".

REFERENCES

1. ACKOFF RL (1978) *The Art of Problem Solving, Accompanied by Ackoff's Fables.* John Wiley, New York, USA.
2. EILON S (1979) Ackoff's fables. *Omega* 7(2), 89–99.

Graduate School of Business MARTIN K STARR
401 Uris Hall
Columbia University
New York
NY 10027
USA

1. 7
79

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 493 to 502
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0493$02.00/0

Energy Forecasting[1]

ANDREW STRATTON

Imperial Chemical Industries Ltd, London

(*Received April* 1979)

A selective and critical review of the state of the art in Energy Forecasting as perceived in assessing future interactions between fuel and feedstock availability for the chemical industry. Models are required of Energy supply and demand and their interaction through the energy market which is subject to national and international policies and conflict and to environmental restrictions and legislation. Significant advances have been made in modelling supply but neither an economic nor a technological approach alone is adequate for modelling demand. In particular there is lack of an adequate model of the industrial sector that reflects the diversity of the production function and adequately incorporates the relationship between technological change, capital investment and energy usage. In the UK the structure of industrial models is over constrained by the aggregation-structure of the national energy statistics. The economic and political forces required to bring about major changes in the capital infra structure are often ignored. There is need to establish logical lower practical limits to energy usage, to develop a system logic structure that provides transfer parameters between various levels and type of aggregation, to develop models which allow probability distributing of variables to be included and to model more effectively the interactions across the boundaries of economics, social and political science, physics, chemistry and mathematics.

INTRODUCTION

THE FORECASTING of expected future demand for a fuel[2] has long been an essential factor in investment decisions by fuel producers, whether public or private owned. In an era of continuing increase in fuel consumption and decreasing cost of oil (by a factor of 2.6 in real terms from 1950 to 1970) forecasting was a relatively simple process of projecting an increase in demand from historical trends and expected future market penetration. The consequences of over estimation of demand were not serious, being a lag in obtaining maximum capital utilisation that was small relative to plant lifetime. The near sixfold increase, in real terms, of the price of oil from 1970 to 1977 and the perceived prospect of further increase has induced major changes in attitude towards energy demand and supply and with it extension of the scope and volume of energy forecasting and the wide publication and debate of the forecasts. Like the debate on 'Limits to Growth' there is polarisation into 'prophets of doom' and 'technological and economic optimism' with frequent and conflicting public pronouncement.

Energy forecasting has become a multi-person game played according to a variety of rules—often not specified—and using models that range from the 'back of the envelope' to highly sophisticated and complex computer simulations incorporating assumptions that, even if fully understood by model designers, may not be communicated to the user or are liable to be overlooked in use. The literature is extensive and this paper does not attempt a complete review of the available techniques and theoretical approaches, but focuses selectively on certain deficiencies perceived by the author in assessing future interactions between fuel and feedstock availability for the chemical industry.

[1] A paper prepared for presentation and discussion at the Third International Research Conference on Operational Research and Management Science, organised by the University of Sussex and held at Bowness-on-Windemere, April 3–6, 1979.

[2] For simplicity in terminology electricity is classed as a fuel.

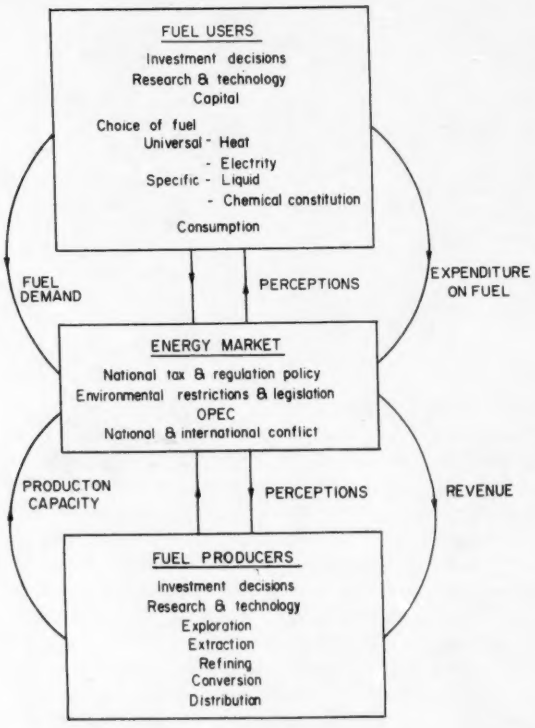


FIG. 1. *The energy system.*

DESCRIPTION OF THE PROBLEM AREA

The real life problem area is depicted in the model of Fig. 1.

At any point in time the fuel user makes decisions on investment in fuel using capital. This decision may be restricted to a specific fuel e.g. a liquid fuel for transport or fuel of a particular chemical composition as a chemical feedstock, or may be a more universal requirement for heat or power, that could be provided by one of a number of fuels. In the former case the decision variables are mainly capital investment, fuel consumption and price[3] whereas in the latter case the decision variables will include, through the choice of fuel, the perceived future relative fuel costs and consumptions, convenience, security of supply and environmental pollution.

The fuel consumption of available capital equipment will depend on prior decisions made by the producer, in which fuel consumption will have been only one of the decision variables. The emphasis given by the capital equipment producer to economy in fuel consumption will depend on his perception of the future significance to the consumer as perceived at the earlier time of research, development and design for production.

The amount of fuel consumed by the user may fluctuate diurnally and seasonally and, to a varying degree, is sensitive to fuel price.

The fuel producer, based on an earlier perception of future demand and return on investment, has made decisions on deployment of resources on exploration and on capital equipment for extracting, refining and/or converting the fuel into the form required by the user. The production capacity has to be adequate to meet fluctuations in demand. Still earlier investment in research and technology will have determined the capital investment options open to the producer.

The interaction between user and supplier is through the energy market, that for oil is international and for other fuels much less so. This market is subject to national tax policies and regulation, to the policies of a powerful cartel,

[3] Throughout this paper it is implied that costing of options includes labour.

Vol. 7
1979

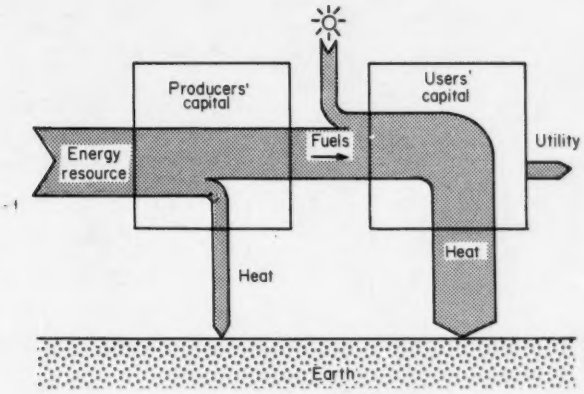


FIG. 2. *Flow of energy through capital.*

OPEC, and to disruption by internal and international conflict and to environmental and other value judgements of groups and organisations that influence policy and legislation.

The objective of energy forecasting is to aid policy decisions on investment in research technology and capital by both user and producer and to aid governments in formulating energy policy.

Some factors that are significant in the system are:

- the interaction between technology and the capital equipment of producer and user that determines the flow of energy necessary to produce the desired utility for the user (Fig. 2)[4]; this flow is constrained by the laws of physics, in particular that the flow of heat and the conversion of heat to work depends on temperature difference[5];
- the availability of skilled manpower for research, development and design;
- the relationship between user utility and economic cost;
- the interactions of the national and international economic markets, including the dynamics of large perturbation;
- the stimuli and constraints generated by group, national, international and cartel behaviour;
- the national and international policies of governments;
- the value placed on national energy or fuel independence.

The model of Fig. 1 may thus be expressed in the alternative form of Fig. 3; components of forecasting models thus range from a descriptive model of group behaviour to experimental and pilot plants.

There are two modes of use of forecasting models, as a source of information or as an interactive tool to explore the effect of decision options. As an information source it is axiomatic that the data supplied is related to specified assumptions of policy, behaviour and constraint and should be expressible in terms of probability distributions conditional on the occurrence of the assumptions. Typical information requirements are:

- energy demand/supply by fuel;
- fuel price and availability;
- effect of fuel price on cost of materials, capital, labour;
- effect of fuel price on future markets;
- timing of any major fuel substitutions;
- timing of any major energy induced material substitutions.

[4] Direct use of solar energy for the user is a redirection of flow by and through the users capital.

[5] Energy 'conservation' is a misnomer; energy is conserved by the laws of nature. The requirement is to produce the desired utility by a minimum flow of heat (Q) and a minimum loss in flow potential, i.e. to minimise the increase in entropy of the system and its environment.

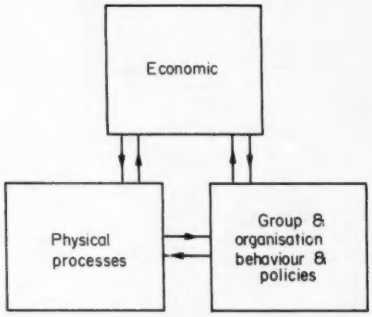


FIG. 3. *Energy system interactions.*

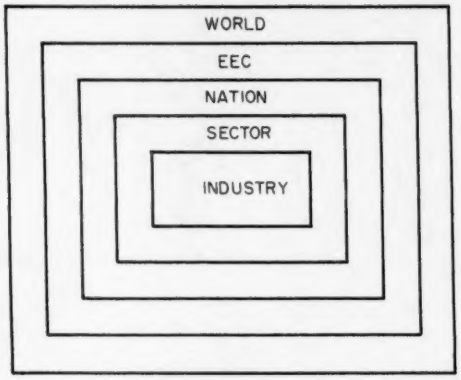


FIG. 4. *System and environment.*

As an interactive tool the model(s) should be usable at any level—industry, sector, nation, group of nations, world (Fig. 4) with the higher levels constituting the environment of the lower.[6]

This is not to say that a large universal model is required.

The interfaces between models at various levels should however be designed to permit the degree of aggregation or disaggregation to be chosen to suit the particular issue being investigated.

AVAILABLE TECHNIQUES AND APPROACHES

It will be convenient to consider these in order:

Energy Supply
Energy Demand
Interaction of supply, demand and policy

Energy supply

The projected future supplies of oil and gas and in particular the domestic production of non OPEC countries that can satisfy a significant proportion of their needs (e.g. US, Canada, UK) is a major factor in energy forecasting for those countries. Expected decline in production is thus linked to the need for projecting the substitution of oil by liquids derived from shale oil, tar sands and coal and substitution of natural gas by gas derived from coal and by direct substitution of oil and gas as a boiler fuel by coal or coal gas.

At the simplest level a descriptive model is used to classify fossil fuel resources in the form of a set diagram such as Fig. 5, with appropriate tonnes inserted, in the main based on experience and judgement; confidence limits on the estimates may or may not be given. The existing industry is based on 'demonstrated economic resources' shown in the top left hand corner, the proportion that is considered recoverable being dependent on economics of 'enhanced oil recovery' techniques; the price of oil that is 'economic' is seldom stated. Recovery rates are often based on maintaining a given number of years reserve. The development of more formal models received a major impetus from the work of Hubbert [9] who showed that in the US Petroleum Industry the discoveries (10^6 BTU/ft drilled) had declined exponentially (1950–70) with cumulative feet drilled. Models have been developed relating the marginal cost of production to cumulative production, of which Fig. 6 is a typical output incorporated into a complete resource–demand model [4], that has been adopted by the US Dept. of Energy. It should be emphasised that these curves are for marginal cost of production and that supply demand elasticity will

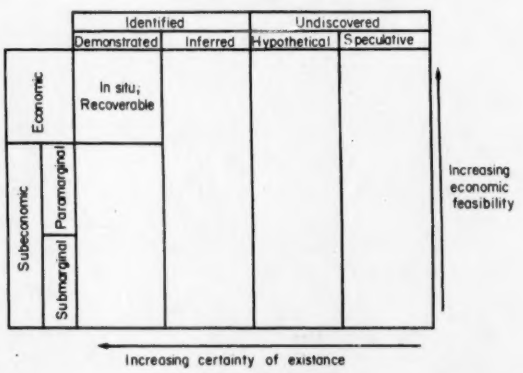


FIG. 5. *Fossil fuel resources.*

[6] This paper focuses on a national level.

Vol. 7
1979

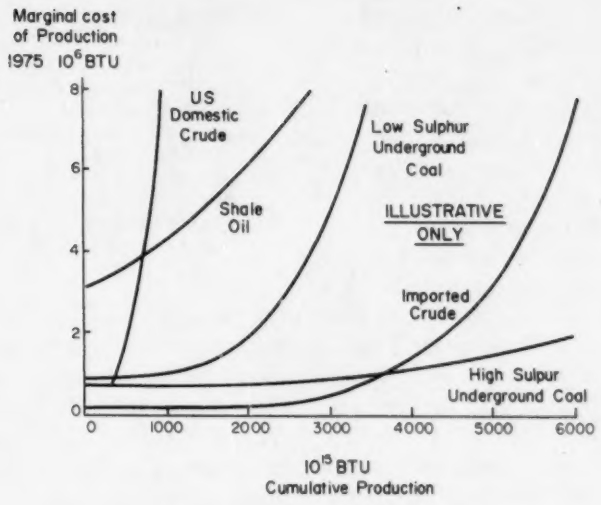


FIG. 6. *Fossil fuel supply curves—US.*

depend on taxation and depletion policies; a policy of constant revenue (as adopted by Norway) will give decreasing production with rising market prices. Another example is [5] that incorporates explicit representations of exploration, reservoir development and production. The discovery history of a region and other geological data are inputs to a statistical analysis of the exploratory process. The resulting estimate of the size distribution of new reservoirs is combined with an evaluation of reservoir economics—taking account of engineering cost, oil price and taxes. The model is applied to the North Sea in [5] and shows a low supply elasticity with the preponderance of North Sea Oil economically recoverable at a $9 price.

The state of the art is summed up by the authors as;

> "Much remains to be done. Yet even in its primitive state this disaggregated method appears to be a significant addition to our existing procedures for forecasting oil supply."

The significant and wide divergencies that occur are illustrated in Fig. 7 which compares the dynamic modelling of Ref. (2) with the 'workshop' approach of group estimation (4). Whilst in reasonable agreement on the AD2000 supply the dynamic response, on the one hand continues the post 73 down trend, and on the other reverts initially to continuing growth.

For substitution fuels, particularly coal derived, the cost estimation is dominated by esti-

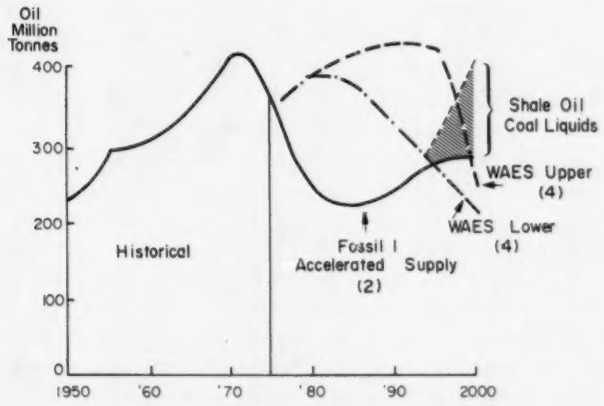


FIG. 7. *Comparison of US oil production forecasts.*

mates of conversion efficiency and capital cost and is sensitive to the mix of produts; conversion efficiencies are often grossly overestimated [15]. For many new processes information is only beginning to become available from small scale demonstration plants. Comparative studies and assessments, conducted by teams of engineers, analysts and economists on a multi-client subscription basis is a major industry in the US. The transition from natural oil to substitute is clearly dependent on the 'break-even' price between the two. This 'break-even' is showing a marked tendency to recede [7] with an increasing cost ratio between the substitute and natural product. Various theories are advanced from more realistic estimates of capital cost to a general inflationary effect of increasing oil price. There is a complete lack of an adequate model for forecasting the differential cost movement of plant cost relative to energy costs and other factors.

Similarly assessment of nuclear power and 'renewable' resources based on sun energy collected by natural or artificial means is dominated by estimation of capital cost and confidence in competing technologies at very different stages of experience.

Demand

The approaches to demand estimation may be broadly classified into 'economic' and 'physical'. In the former, attempts are made 'top down' to derive relationships between economic variables and the use of energy and to use these as a basis for extrapolation into the future; in the latter, estimates are made 'bottom up' of the amount of energy required for particular usages and postulates are then made on how this will change in the future.

The simplest economic attempts at demand estimation link growth in energy consumption (E) to growth in GNP or GDP (P) through the elasticity

$$\epsilon = \frac{P}{dP}\frac{dE}{E};$$

even if the elasticity were to exhibit consistent behaviour it is argued that it is too gross to reflect the many interactions within demand [3]. The position on individual elasticities for fuel demands and cross elasticities for substitution is summarised in the Report of the Working Group on Energy Elasticities [13], that gives 59 references;

> "From the available data it is not possible to derive a full set of UK energy elasticities with useful accuracy. Estimates of future patterns of fuel consumption have therefore to rely on a variety of techniques adapted to the information available in each area on factors such as supply constraints, economic and political objectives and taking account of general technological information and judgements about, for example, the rate and extent to which consumers would respond to large energy price rises".

On the other hand it is argued that an index of the ratio of total energy consumption to GNP for the US between 1880 and 1975 reveals three distinct periods first of increase (1880–1920) then of decrease 1920–45 and since 1945 no persistent up or down trend, albeit with fluctuations [14] and that consequently major forces will be required to depart significantly from this constancy. The author argues inter alia that greater attention has to be paid to engineering data and comparative international practices which set forth the range of feasible technological alternatives that might become economically attractive as energy prices rise and criticises the assumption in most econometric energy models that the conditions under which energy is supplied in the future will not impede future productivity growth.

> "Both capital formation and productivity are taken to be unrelated to energy development in most models."

In the sectoral approach to demand estimation used by the UK Dept of Energy [17] and other workers [3] the division adopted is shown in perspective in Table 1 which gives the 1977 consumption.

In the 'physical' approach, domestic demand is separated into space heating and electrical appliances (ownership approaching saturation) and the estimation of reduction in heating demand through insulation which requires the response time of the householder to be modelled. Commercial and Public Administration requires estimates of heating and lighting and electrical equipment; direct consumption by agriculture is very small.

Modelling Transport starts from estimated travel demands, road and air dominating; growth in number of cars per household and estimates of improvement in consumption. Longer term studies have been made of the possible substitution by electric vehicles [6].

Vol. 7
1979

TABLE 1. UK ENERGY CONSUMPTION 1977

Energy delivered	10^{18}J
Domestic	1.59
Commercial public administration and agriculture	0.80
Transport	1.38
Iron and steel	0.52
Other industry (inc. chemical feedstock)	2.17
Oils, etc.	0.19
	6.65
Losses in energy industry	2.64
Gross Primary Energy	9.29

Turning to industrial energy demand the need for a disaggregated approach is illustrated in Fig. 8 which shows, by sector, UK energy usage per £ added value. Added value is selected as an index, being both the constituent component of GDP and a measure of the combined cost of labour and capital. (See Appendix for note on the UK industrial Energy Statistics).

In the 'economic' approach, energy is of course only one of the factors of production. There is considerable attention in economics literature to generalised production functions that link the volume of production to the capital and labour employed through log–log functions, such as the Cobb–Douglas [12]. With the growing interest in energy as a factor of production, production functions are being investigated that include energy, disaggregated into individual fuels [8]. There is much discussion whether energy and capital are complementary or substitutable (i.e. energy can be saved at the expense of more capital); reference [8] indicates that the relationship between energy and non-energy inputs is predominantly that of substitution. Whatever relationship has obtained in the past, however, there is no immutable physical law that this relation must obtain in the future and some recent work in the chemical engineering field indicates that, by appropriate thermodynamic design, both capital and energy can be saved [11].

Reference [3] starts with a disaggregated approach to industrial energy usage comparing

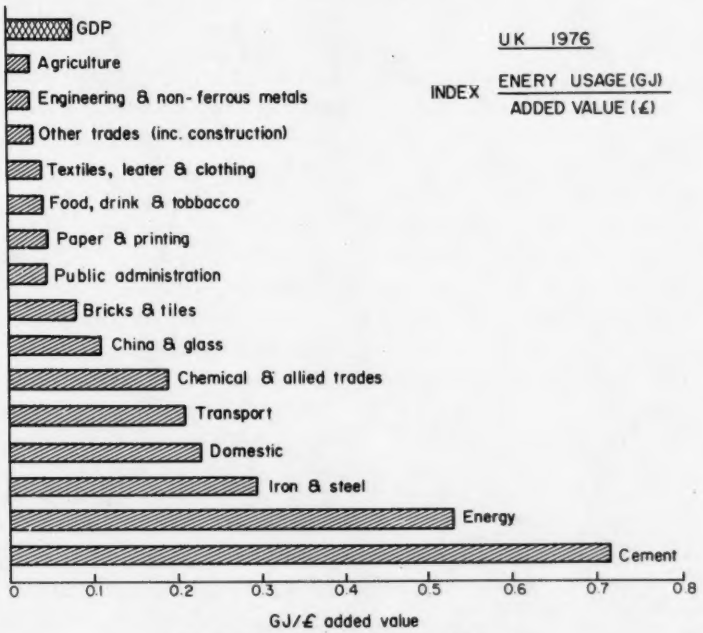


FIG. 8. *UK energy usage for £ added value.*

historical trends in energy usage (heat supplied) with unit of output of each manufacturing industry but then deduces that:

> "there is no way of accurately predicting changes in the composition of manufacturing output over two decades nor technical change which could result in shifts in energy consumption between the various sub-sectors."

Reference [3] then reverts to assumptions about the growth of manufacturing production and the elasticity of energy usage in relation to output.

There is, to the author's knowledge, no model of the industrial sector that incorporates adequately,

- the diversity of energy labour and capital usage that is evident in Fig. 8;
- the relationship between technical change, capital investment and energy usage;
- practical physical limitations to reduction in energy usage[7].

Recent informal enquiries at a senior level in the US Department of Energy brought the response that, compared with resource modelling, there is a major deficit in capability to model the industrial sector that could take some five years to fill.

Interaction of supply, demand and policy

At the simplest level of energy forecasting, no account is taken of the interaction between supply and demand nor of the political factors that would be involved in implementation. For example, during the writing of this paper Reference [10] was published. This is a very detailed sectoral analysis of energy usage, leading to estimates of reduction in energy usage that should be possible by technological change; a very large number of assumptions on realising potential savings are involved, which are based on numerous illustrative examples. It is then assumed that the numerous investment decisions will be made by individuals, and industries in a timely manner, without consideration of the economic and political forces required to bring about this transition. The magnitude of these forces that would be required may be judged from the conclusion that a doubling and tripling of the UK GDP by 2025 can be accommodated within an energy consumption that is less than or about equal to the 1973 consumption.

By contrast, Reference [16], which develops the same thesis (for the USA) that energy conservation can bring about a stabilisation in energy demand with rising GDP, compares the capital investment for conservation with that for increasing energy supply and then declares:

> "Our proposed policy for accelerated energy conservation depends heavily upon mandatory measures to improve end-use efficiency. This approach will inevitably raise arguments against tampering with the so-called 'free market'. Direct intervention must be considered, however, because price alone cannot provide sufficiently strong motivation for accelerated conservation."

This interaction between open market forces and policy is incorporated formally into the model of Reference [4]. In the demand part of the model energy consumers make decisions to utilise oil, gas, coal and electricity based on both price and availability of the fuels; energy producers, in turn, choose to invest in the product that maximises the industry rate of return subject to environmental constraints and market limitations.

The interaction is dynamic; as conventional resources are depleted the costs and prices of oil and gas rise. On the demand side, price increases encourage consumers to conserve energy and shift away from oil and gas towards coal and electricity. On the supply side, rising conventional production costs stimulate domestic oil and gas producers to develop alternative sources of oil and gas.

The effect of government energy policies (such as price de-regulation, research and development or environmental legislation) can be simulated and the model user can alter the model's decision rules to match government policies. (The model predicts that even with vigorous US Government energy policies the

[7] The term 'useful energy' is often used, the ratio of 'useful energy' to 'energy supplied' being efficiency of usage. Whilst the second law of thermodynamics enables a limiting efficiency to be defined for a heat engine (Carnot cycle), a reversible process with zero increase in entropy can only proceed infinitely slowly. Any practical process must produce an increase in entropy of the system plus environment, that is related to the rate at which the process takes place and thus sets a practical upper limit to efficiency of energy usage. There are thus many cases in which 'efficiency' has no well defined meaning. The thermodynamics of systems operating at finite rates is only recently beginning to receive attention [1].

Vol. 7
1979

future situation is much more pessimistic than most energy studies allow.)

The model is essentially deterministic and the author is not aware of any model that incorporates probability distributions of variables.

SUGGESTIONS FOR FURTHER DEVELOPMENT

A high degree of disaggregation in modelling demand appears inevitable and there is particular need to improve our capability to model the industrial sector. In the UK the structure of models of the industrial sector is over-constrained by the aggregation and structure of the national energy statistics (Appendix) and in particular the following need to be individually identified;

- sectors producing basic materials (metals, plastics, wood, paper, cotton, wool, glass etc);
- components of the engineering group;
- the construction industry.

This would enable a restructuring in which the total energy requirements to produce the materials used by the economy and to provide the capital for manufacture, material and energy conversion can be clearly identified.

An important area for further development, particularly for modelling the industrial sector, is the relationship between increase in entropy, energy usage, capital, and process rate to establish logical lower practical limits to energy usage.

A particularly important 'macro' problem is the development of a system logic structure that provides transfer parameters between various levels and type of aggregation. For example, the interface between an 'economic' and a 'physical' model can still be familiar parameters such as elasticities, which are no longer historically derived but generated in another model. Similarly input–output matrices can be used but with variable coefficients derived from a physical industry production model.

The need for development of a model which allows probability distributions of variables to be included has previously been mentioned; this is a fundamental deficiency in our ability to model large stochastic systems.

In conclusion, energy forecasting is a typical large 'messy' problem wherein:

- Relatively small differences in projections of supply and demand over the time scale of implementation represent major differences in investment policy;
- physical laws and technology constrain the available options but individual and organisational behaviour determine the implementation;
- sets of individually reasonable assumptions can be chosen that support a wide range of different forecasts.

An approach solely from economic modelling depends on elasticities that, even if determinable are functions of the existing capital infra structure and existing individual and organisational values. An approach solely from physical analysis, and in particular reductions in energy usage, whilst it may define potential new capital infra structures, cannot determine the time scales and dynamics of implementation nor the changes in values and policies that would be necessary to bring about implementation.

We have thus returned to the starting point with the need for the future to model more effectively the interactions of Fig. 3 across the boundaries of economics, social and political science, physics, chemistry and mathematics.

ACKNOWLEDGEMENT

This paper is based on investigations undertaken by the author as a member of ICI Limited; the views expressed are personal and are not necessarily those of the Company.

REFERENCES

1. Anderesen B, Salamon P & Berry RS (1977) Thermodynamics in finite time; extremals for imperfect heat engines. *Jl Chem. Phys.* **66**, 1571–1577.
2. Carrol LW (1977) *Report of the Workshop on Alternative Energy Strategies.* McGraw-Hill, New York, USA.
3. Chesshire JH & Surrey AJ (1978) Estimating UK Energy Demand for the year 2000; a Sectoral Approach. *SPRU, University of Sussex Occas. Pap. Ser. No.* **5**, February.

4. DARTMOUTH SYSTEM DYNAMICS GROUP (1977) Fossil **1**, New Hampshire, USA.
5. ECKBO PL, JACOBY HD & SMITH JL (1978) Oil supply forecasting: a disaggregated process approach. *Bell Jl Econ.* **9**(1).
6. ENERGY FOR TRANSPORT (1978) Long Term Possibilities. *Energy Pap.* **26**, HMSO, London, UK.
7. 'ESCOE' ECHO (1978) Newletter of the Engineering Societies Commission on Energy Inc. **2**, 21st October.
8. HALVORSEN R & FORD J (1978) Substitution among energy, capital and labour inputs on US manufacturing. In *Advances in the Economics of Energy and Resources* (Ed. PINDYATA RS), Vol. 1, JAI Press.
9. HUBBERT MK (1974) US Energy Resources. a Review as of 1972. *US Senate Committee on Interior and Insular Affairs, Ser. No.* **90–40 (92–75).**
10. LEACH G ET AL (1979) A low energy strategy for the United Kingdom. *Int. Inst. Envir. & Devt.*
11. LINNHOFF B & FLOWER JR (1978) Synthesis of heat exchange networks. *AIChE Jl*, **24**(4), 633–654.
12. OTT DJ, OTT AF & YOO JH (1978) *Macroeconomic Theory*. McGraw-Hill, New York, USA.
13. HMSO (1977) Report of the Working Group on Energy Elasticities. *Energy Pap.* **17.** HMSO, London, UK.
14. SCHURR SH (1978) Energy, economic growth and human welfare. *EPRI Jl.*
15. STRATTON A (1978) Fuels and chemicals from coal—some energy and resource implications for the UK. *Chem. & Ind.* 551–560.
16. WIDMER TF & GYFTOPOULOS EP (1977) Energy conservation and a healthy economy. *Technol. Rev.*
17. DEPARTMENT OF ENERGY (1978) Energy Paper **29**, HMSO, London, UK.

ADDRESS FOR CORRESPONDENCE: *Dr Andrew Stratton, Imperial Chemical Industries Limited, Imperial Chemical House, Millbank, London SW1P 3JF, UK.*

APPENDIX

Statistical data

Official publications in the UK contain massive detail on the conomy down to annual gross sales, added value, number of employees, wages and salaries, investment in various types of capital for each Order in the Standard Insustrial classification (SIC) and nearly every Minimum List Heading (MLH) into which the classifications are divided [9]. The Digest of UK Energy Statistics groups industrial energy consumption as follows:

Iron and Steel	Order VI MLH 311
Engineering and other metal trades	Order VII to XII inclusive approx MLH 321, 22, 23
Food, Drinks and Tobacco	Order III
Chemicals and Allied Trades	Order V plus MLH 263
Textiles, Leather and Clothing	Order XIII, XIV, XV
Paper, Printing and Stationery	Order XVIII
Bricks, Tiles, Fireclay and other Building material	Order XVI MLH 461 and 469 only
China, Earthenware and glass	Order XVI MLH 462 and 463 only
Cement	Order XVI MLH 464
Other trades	Order II excluding MLH 101, 104 Order XVII, XIX, XX Order XXI MLH 603 only; MLH 892

Reference [9] uses the following classification:

Energy Industries	Order II MLH 101, 104 only Order IV less MLH 263 Order XXI MLH 601, 602 only

The data on Order and MLH used in the Energy Digest is not given in official publications; the above list was obtained verbally from the Dept of Energy.

Inconsistencies occur in many published papers in the fuels and feedstock used by the chemical industry. In Energy Digest petroleum feedstock (naphtha) is listed as part of 'non energy' use and is not included in the above listing for the chemical industry. Natural gas used as feedstock is, however, included.

A major defect in Energy Statistics is the high level and type of aggregation used wherein:

- energy intensive material production e.g. aluminium and other metals, paper, is included with the less energy intensive production industries
- the construction industry is lumped with "other trades"
- a wide diversity of industries is aggregated under "engineering"

Reference [9] Business Monitor Census of Production PA1000 HMSO.

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 503 to 512
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0503$02.00/0

Personal and Organisational Value Systems: How Should we Treat these in OR Studies?

KC BOWEN

Defence Operational Analysis Establishment, Ministry of Defence, UK

(*Received May*, 1979)

Values that are held by individuals, or jointly-held values that are attributed to organisations, affect all stages of the process of decision. It is often difficult to determine how and whether the value systems of decision-makers and analysts should affect the analysis that is carried out; to know what these values are and how to measure or describe them; and to put numbers to values, or the consequences of values, if these are to be represented in models. This paper does not attempt to define the way in which value systems are created or used. It does not deal with value systems as they are treated by decision analysis and theory, both because these ideas are well-documented, and because it seems to me that the attribution of utilities and subjective probabilities, and an appropriate axiom structure, may be impossible or misleading. Arguments for a different approach are given in Part 1, and it is suggested that something less prescriptive and less precise is needed. Part 2 looks at some of the more hopeful ways in which this is being, or might be, attempted. The philosophy generally is to help decision-makers to use their own values with more insight, rather than for these to be defined and used by the analysts. Directions in which further research should be encouraged are thereby implied. Part 3 looks at a number of associated ideas which might well be seen as further material for debate. The paper concludes with these, but it does not draw conclusions. It is presented as a basis for discussion and is necessarily biassed by my own values.

PART 1—INTRODUCTION AND INTENTION FOR DISCUSSION

Values and beliefs

IF WE ARE to get good measures of behavioural variables, we need good operational definitions of the terms used to describe them. We do not seem to have such definitions of *value* and *belief*: definitions that might serve are by no means agreed. The treatment, by Ackoff and Emery [1], of these and other concepts makes it clear that the context of the choice situation affects any absolute statements, including measures, that might be made. We can always ask people about their values, but, in the end, we can only infer what values they appear to hold by analysing their behaviour, including their statements, in a number of situations. In any one of these, however, there are likely to be conflicting beliefs and values, and the compromises made may not have any obvious basis. It is commonplace that 'people's actions belie their beliefs'.

In a draft for a forthcoming book, *Subjective Worlds in Policy and Organization*, Colin Eden, Sue Jones and David Sims discuss values under the provocative chapter-title of 'Beliefs about Values'. They do not find any depth of understanding of individual or organisational value systems in the literature, nor do they pretend that they have yet achieved such understanding for their own purposes. In setting down my own thoughts as a guide for discussion, I offer no more than a statement of what I think I am talking about.

Values and beliefs are vague and shifting bases for guiding choice (although in certain contexts, beliefs might be held to be deeper and to change only slowly). They are personal rules, sometimes shared with others. They are created

[1] A Paper prepared for presentation and discussion at the Third International Research Conference on Operational Research and Management Science, organised by the University of Sussex and held at Bowness-on-Windermere, April 3–6, 1979.

from data accumulated over a lifetime. They are seldom made explicit, certainly never fully explicit, and, even when relevant, they are not necessarily used. The failure to use relevant rules may be because the situation is not clear, because there has not been time to process data, or because the set of available rules are neither complete nor mutually consistent for the particular choice-situation faced. Probably all of these reasons operate together to some degree. It is perhaps as a consequence of these reasons that the words 'value' and 'belief' may conjure up different concepts at different times and for different people.

At the simplest level of appreciation of what *value systems* and *belief systems* may comprise, we could consider three distinct concepts. Values might be seen as a sub-set of beliefs, elemental beliefs which enable more elaborate beliefs to be formed. Alternatively, beliefs might be seen as a sub-set of values, that is, values of a special kind. The third concept would treat beliefs and values as separated sets that interact strongly: values might be formed to accommodate better a belief, or beliefs might be formed as a consequence of values. I can see no reason why any of these concepts should be thought to be 'wrong'. I am prepared to regard belief and value systems as very fuzzy sets, with ideas in the universe having some degree of membership of each.

This does not imply that I think that a fuzzy calculus should immediately be applied, so as to bring values into a category compatible with the measurement-categories of statements about the nature of physical characteristics and events. Indeed, we do not yet know how to do this. But in seeking ways of dealing with values, it would be wise to recognise that we may be unable to progress further than fuzzy measurements in our search for OR (Operational Research) methods of incorporating measures of value, at least in the forseeable future. We may not even get that far.

Utility and value

The concept of *utility*, something which measures the usefulness of possessing something or doing something, is far from simple. Also, as a word, it may confuse. I observe that it is not used in either of the two major books by White [45, 46] on decision, nor by Ackoff and Emery [1] except as a special case of *intention*. On the other hand, value, except in a more or less financial sense, does not seem to be a word much used by Lindley [32].

Sometimes, utility and value are, in the technical sense, used synonymously; in such cases, there is an implicit concept of an absolute usefulness and worth, the measurement of which is an ideal in Ackoff's sense [1]. Such a measurement would stem, in theory, from an assessment of the absolute merit of some choice in a defined situation, and there is an underlying assumption that some true definition and measurement of value is achievable and is known to and accepted by those making and using the assessment. In practice, we can only move towards such an ideal of measurement: any so-called utility measure can be no more than a surrogate for value in the ideal sense. However, in some contexts, value will have, or will be understood to have, a more pragmatic meaning, something akin to market value. It may then be seen as a useful surrogate for utility and indeed may be treated as if it were a true measure of utility. There is no doubt that concepts such as utility and value are necessary and can themselves have utility and value; but, inevitably, the way the words are used raises both semantic and operational problems.

Decision-theory is important and I do not doubt its considerable conceptual value. However, decision-theoretic approaches, based on, for example, maximum expected utility theory, do not satisfy me as practical tools for general use. The measures required by theory not only have to be ideals (not surrogate measures), but they have to be special measures obeying a logically complete set of axioms on preferences (implying value judgements). Theoretically, it is not possible for an individual's action to be in accord with his 'real intentions' unless he maximises expected utility. Only by denying the impossibility of attaining ideals, can I empathise at all with a *cri-de-coeur* of Lindley when replying to a paper by Tversky and Suppes [33].

> "There is one comment that I would like to address to psychologists present. Why do you spend your time studying how people make decisions, when we know how they should make decisions? Would it not be better to devote your energies to teaching them the principles of maximum expected utility?"

In part, Lindley's advice has been taken. Kaplan and Schwarz [28, 29] have put

Vol. 7
1979

together two volumes of papers on human judgement and decision processes which include ones which treat applications of decision theory—they also draw attention to new approaches by psychologists which operational researchers might well borrow; in particular, Social Judgement Theory, which aims to help people to understand better their own necessarily subjective processes (chapters by Hammond, KR *et al.* in [28, 29]).

Three chapters in [28] are of particular interest in the context of decision theory. Gardiner and Edwards treat multi-attribute utility measurements in such a simple and direct way that the mechanical nature of what they are doing is emphasised. Ironically they conclude: "We know of no public context in which even limited experiment with the methods we advocate is occurring". Coombs purports to show that subjectively expected utility theory does not reflect human behaviour. He tests what he calls 'portfolio theory', which assumes a single-peaked preference function over risk. He concludes that his is a more satisfactory descriptive theory. One can, of course, argue that risk is thereby shown to have utility for some people so that the earlier numbers, treated as utilities, were poor surrogates. Shanteau, likewise, studying an information-integration approach to decision-making, finds that the additive utility hypothesis does not hold, which, by definition, implies that his measures fall short of the ideal.

By the very complexity of judgement and decision, it cannot be surprising if measures of utility fall far short of what is intended. But for most decision analyses, difficulties are compounded by the need to assess probabilities associated with future events. Often when a decision is made, it is *as if* such probabilities had been used. However, the processes used to establish these probabilities, as if they were absolute properties of an individual's values and beliefs, are not easy to accept; fuzzy logic moves away from probabilities for similar reasons. There may be cases in which probabilities associated with value judgements are usefully measurable, but the conditions might be stringent: for example, fairly simple, well-understood problems, with the person taking the decision being also in control of, or in full sympathy with and understanding, the analysis.

A proposed framework for discussion

To debate, over a short time, the issues raised so far, would no doubt be exciting and controversial, but it would not perhaps be profitable. I therefore wish my statements to be taken, initially at least, as indicating why I seek softer scientific approaches to the task of dealing with values within OR studies. The processes I seek are to help people to formulate and understand their own problems more clearly; or to help people in *their* task of making the final choice, when analysis of the decision has gone as far as it can in the task of presenting the pros and cons of alternative choices.

My intention in Part 2 is, therefore, to identify, and describe briefly, processes which seem to fall into the desired categories in some respects. There is almost certainly research of the same general type, perhaps even applications, of which I have not heard. To get a more comprehensive list of such research is one purpose that I have in mind.

Another purpose is to discuss what limitations there are on providing a theoretical basis for such research. The relationship of trust between analyst and client, and the nature of communication and interpretation required, are likely to differ from what is deemed to be normally satisfactory in OR. At a less demanding level, statements of methodology would be valuable: since values pervade all aspects of all studies, such statements should have a wide applicability for improving the practical impacts of less demanding studies, in which OR is already relatively successful.

Also, I believe that we need to discuss whose values and beliefs we are concerned with. In an organisation, *the* decision-maker is a somewhat unreal concept. There may be a responsibility-taker (an idea borrowed from Professor M Shubik of Yale University). His values are important but perhaps no more important than the values of others who take part in, and affect, the decision-process. These others include some who are outside the organisation, and in particular, they include the analysts who may be inside or outside.

The concept of the values of an organisation are complex. I hope that we can accept this complexity as resulting from the various interactions between individual values, without going too deeply into semantic argument on

concepts such as *power*. I accept however that the interactions have to be discussed and understood to some extent, so as to be able to perceive which methods acknowledge adequately the part played by values in organisations. Some of the larger-scale societal problems, for which OR should aspire to offer resolutions, include perhaps the most difficult 'organisation values' that can be conceived. If it can be said that there are methods in the making that address such value-ridden problems, even the more theoretical and conservative among operational researchers may concede them a place within OR.

PART 2. SOME NEW APPROACHES TO HANDLING VALUES

There are a number of topics, stemming from recent research and application, which are concerned directly or indirectly with value systems and which, I believe, merit more attention than they have been given to date. They include analysis of options; hypergame theory; structural mapping, personal construct theory; fuzzy sets; and a number of ideas stemming from conflict research, research gaming, and thoughts about the general role of the analyst.

I shall not deal with these individually, but introduce them under more generic headings, since personal and organisational values enter into all phases of the total decision process along with many other, equally elusive, variables.

Problem-formulation

Understanding someone else's problem is difficult. We work hard at the task of communicating with those responsible for decisions, but we have no clear guidelines, certainly no explicit ones, that tell us what to do and what not to do. Generally, there are many cooperating or conflicting groups (or individuals) concerned with 'the' problem. Not all will perceive it in the same way.

Eden [17, 18] has developed a problem-formulation procedure which has something in common with counselling. He uses ideas [3, 4, 31] from cognitive psychology and sociology (Kelly's work [31] is directed towards clinical diagnosis and psychotherapy). Eden's purpose is to help those who own the problem to develop their own perceptions through constructs (which define an outcome) and the relationships between them. He uses directed graphs to display these ideas visually and a computer programme to enable questions to be asked about the resulting complex, and indirect, associations. The language of the client is used throughout. This is practical work with real decision-makers. It has been successful in several directions.

It has made people within the same organisations aware of their different purposes and goal structures, where they had hitherto assumed a congruence of values. It has promoted insights into the nature of the organisation and how it might develop in order to improve its capability. This latter aspect seems very important. Even if there is something approximating to a common set of values in an organisation, a lack of appreciation of how organisational structure limits the execution of an agreed policy implies a lack of appropriate constraints on options for decision.

There is an intriguing problem of values and beliefs inherent in the communication of the methodology of such studies. There are, of course, restrictions on what can be said because of confidentiality, but there are also difficulties in making explicit the actual, and subjective, process by which the appropriate, and simple, techniques are actually used. It is not easy to infer from material so far published, exactly how the counselling role is carried out: observation of the work by nonmembers of the analysis team would be an infringement of privacy. It may therefore appear that nothing other than a somewhat more formalised communication between analyst and client is going on, and the natural response is 'we all do that sort of thing'. Similar problems are evident when one reads of Burton's methods [13] of helping to resolve conflict situations, where, again, all depends on who the counsellors are and precisely how they operate. Partly at least for these reasons, Eden's work is not yet recognised as being in the mainstream of OR, nor is it easy to see how to measure its usefulness, other than by expressing a belief, as I do, based on empathy achieved through many discussions and on my own values as expressed here and elsewhere.

Indeed, my own studies of conflict, and those of colleagues at the University of Sussex, have followed lines not inconsistent with those of

Vol. 7
1979

Eden, although the approach has been more theoretical. Dando and Sharp [15, 16] have discussed some of the attitudes that restrict the view of what OR is and what it can do, and Smith [42] has written in support of their views (interestingly he prefers "to view OR as a form of organisational counselling"). I have used a diagrammatic 'language', developed in earlier conflict studies [9], to examine problem formulation [7, 8, 11], albeit, so far, in minor ways: it is certainly possible to get insights into some of the complexities of problems and of different perceptions of them, and it may be a step towards a check-list of what questions might be asked and what broad procedures might be followed. Because my approach necessarily has to define the elements of the problems and their relationships with attention to perceptions of the problem, it may well turn out to have close links with the repertory and implication grid methods used by Eden (see Bannister and Fransella [3].

I know of no other formal processes, in use or under development, which address the key issues of problem-formulation: how the analysis and decision-making organisation works; how the decision-making may be constrained—by events, organisational structure and differing value judgements; and how much of the environment, in which the system to be changed operates, should be taken into account. I believe that explicit models are essential, although 'model' here does not necessarily imply more than a logically structured description: it certainly does not imply a mainly mathematical or numerical treatment.

Model-choice and model analysis

Smith [42] draws attention to the dangers of the analyst's values intruding into the modelling and so affecting, often in subtle or barely discernible ways, the advice proffered. (I note, in this connection, that Eden and his colleagues use the same techniques to help them understand their own problems in the primary counselling task.) I accept that, in some measure, the values and beliefs of the analysts are bound to affect their secondary decisions (as White calls them [46]) and so react on their contribution to the primary decision. But there is very little attention paid to this, and it is as if they assume, with their general background of mathematics or the physical sciences, that their values are somehow objective and rational. Certainly there is little in the form of techniques to aid analysts with the secondary decisions.

In the literature of statistics and OR, I have found little guidance as to what conditions should be satisfied before a particular technique is used. I have explored Games in this way [12], in order to say when a game should be used, and what sort of a game it should be. I note here that games in some form are always needed if human judgement is to be introduced into models. Here and there, good accounts can be found of the conditions that must hold if statistical models and procedures are to be valid, and I have recently seen an attempt to deal with the assumptions underlying the application of mathematical programming models. But these are exceptions and they cover only a small area of OR modelling.

What is required is a classification, for all sorts of models, which links data, assumptions and constraints, methods of use, and interpretations of output. Nowhere is this more necessary than in the field of computer simulation. Such models are difficult to justify, except as a last resort, because they are hard to control and analyse. Their use as a learning tool (a game for the analyst) has recently been discussed, by several individuals and groups quite independently; the idea is to be able, with confidence, to develop and use cheaper and more understandable models.

Even when we have decided on a model, what analysis to carry out is hard to decide. Some such decisions will, of course, be part of model-choice. Others will depend on the limited reality of the model, and the fact that the dependent variable is only a proxy for some concept of goodness of the new system. The sort of analysis and presentation that will be meaningful and satisfying to those the analyst is helping becomes of some importance, since their value systems and that of the analyst may well conflict.

I believe that, for the tasks that are peculiarly those of the analyst, there should be a careful check on the ways in which values and beliefs enter a supposedly objective appraisal. If this is done, there should be a greater awareness of the more profound behavioural issues inherent in the stages of problem-formulation, of the communication and use of the findings

of analysis, and of the ultimate implementation of changes to the systems studied. It is to the last two of these stages that I now turn.

Communication and use of the findings of analysis

I have always maintained that the correct end-point for an operational research study is the displaying of options with measures or statements of the criteria on which choice depends. I do not see it as necessary, or proper, to attempt to machine the decision. There may be criteria not included in the study (despite all efforts), the data may be uncertain, and confidence in the measures will generally be very subjective. Nevertheless, decision-makers should not be left with an intractable basis for choice, and they may welcome advice as to how to set about the task which is properly theirs. I do not, at present, think that decision theory and decision analysis offer acceptable solutions. They require decision-makers to agree that when certain of their values and beliefs are pinned down, generally in numerical form, a simple probability logic (or, in the case of maximum expected utility, a not-so-simple logic) will represent the proper process of choice. Although interest has been shown in many places, I cannot find case-histories of a decision on a major problem having been taken or strongly influenced, by such methods. It may well be that the methods might have importance as a learning process for the decision-maker, but they are not offered in suitable form for this purpose.

Simpler methods have been used. Maximin solutions (which are pessimistic) have been found acceptable in some military analyses [27]. Minimax regret has been a useful mind-clearer in the first round of discussions on other military studies. Both, of course, involve the concept of utility as a common metric.

What can we do if there is no common metric? In order to make the decision-maker's task bearable, we may have deliberately left out many options and criteria, or merged them, in ways that we cannot support by analysis. Rivett [38, 39] has suggested the use of multi-dimensional scaling techniques to map an individual's preferences, based on very meagre information about his indifferences between choices, when the table of options and criteria is large. The research at present concentrates on the robustness of the preference orderings produced [14], and early results are encouraging. There has been a positive reaction from Green [24], which I welcome because the ideas put forward by Rivett are of great potential importance. Indeed, I know of no other approach which offers a structured way for a decision-maker to rank such preferences without asking him to describe his values numerically and without demanding that he be precise when he would prefer to be vague. Fuzzy sets ought to help but applications as yet seem to be almost entirely in the area of control of mechanical systems.

'The' decision-maker is a convenient fiction. In practice, as I have already discussed under problem-formulation, there are many people involved, and they do not have common values and beliefs. Radford's Analysis of Options [37], a development from the theory of games and metagames, examines the possible results of the various feasible interactions and the ways in which communication, negotiation and bargaining can take place to produce a resolution that is stable (in the sense that deviation from it is not ultimately advantageous to anyone). It is of relevance, of course, for any individual decision-maker, to help him come to an understanding of the problem he faces and the decision-process for which he is jointly responsible. I have included it here, in the context of post-analysis communication, because there has ultimately to be analysis to provide a clear insight into the options available and the preferences among them.

Radford provides an illustrative case-history which is an after-the-event analysis of the Toronto airport decision (1967–75), and it also refers to other conflict descriptions and analyses by Nigel Howard and others. These are all potential uses of the analysis of options. There still seems to be a lack of evidence, in the literature, that the approach has been used and found to be useful by those responsible for decisions. I understand that there is encouraging progress in this direction, and I look forward to Radford's 'Analysis in Support of Strategic Planning' which is expected towards the end of 1979.

I repeat that I have not been able to get documentation to support claims that decision analysis and other decision-theoretic aids have been successfully employed. It certainly seems

Vol. 7
1979

that benefit can be gained by the structured thinking that all such methods offer, preferably if they are used by the decision-makers, but the presence of a suitably trained analyst, who communicates well with his clients across the whole span of the problem and the decision-process, appears to be essential. As with Eden's work, discussed in the section on problem-formulation, the personality and behaviour of the analyst are critical factors.

Within the total coverage of complex decision problems, Radford refers to other techniques, which help in the earlier stages, notably cross-impact gaming and Delphi on which there is a wealth of literature (e.g. Helmer [25, 26]). These are certainly relevant to the data decisions that analysts face in the modelling stage, but may also play a direct role in aiding subjective judgement when objective analysis has gone as far as it can. There is, however, the more recent development by Bennett [5, 6] of hypergames. The intent here is to describe how actions in conflict may appear 'illogical', because the 'game' is being differently perceived by different participants. The research has not yet been tested in application, but it is already clear that it provides a useful approach to the analysis of past events and aids insights into current events. It is compatible with and not an alternative to the analysis of options.

Implementation of change

A lot of what I have already said has been linked with the implementation stage. There are only a few things to add. As a system is changed we need a further analysis. This may tell us something useful about the preceding analysis. We also need analysis for reassessment of prior decisions, when the environment was predicted on less up-to-date knowledge. It would perhaps be useful if earlier analyses could be less detailed, in order to provide less prescriptive and more flexible policies, with the effort saved being devoted to analysis at the moment of actual implementation of change.

It is in this last sense that implementation is a final issue for analysis. In its other aspects—the capability of the organisation to manage change, the desires of the organisation for particular changes, and the conflicts over change within and outside the organisation—implementation factors are issues for problem-formulation.

I am not aware of methods that specifically relate to values and beliefs which affect the ultimate implementation problem, nor am I aware of any practice of follow-up analysis, other than to restart the analysis when what has been done turns out to be ineffectual. It is important to realise that, all too often, much of the ultimate action is based on personal beliefs and value judgements with little aid from analysis, unless those who help to formulate policy remain associated with the implementation process.

PART 3. IN CONCLUSION

There is a great deal of other work which is relevant to the quest for more adequate methodological treatment of values and beliefs. There is potentially relevant material in political science and international affairs studies. In particular, cognitive mapping techniques are used (e.g. Shapiro and Bonham [40]), although this use does not appear to be as advanced in development and application as it is in Eden's work previously referred to. Secondly, there is Hammond's work on application of Social Judgement Theory referred to in Part 1; I have not yet been able to study this in any detail. Thirdly, the problems of reconciling values arises in work on how to measure research output: Freeman [20] summarises most usefully the advantages of doing something, while pointing out the dangers of doing it at all.

On the general question of what operational researchers ought to be concentrating on, Emshoff [19] proposes work predominantly in support of problem formulation, specifically asking for models for the decision-maker to use. De Waele [44] has drawn attention to the relevance of Jung's classification of personality types in discussion of who should do what analysis for whom. Mitroff and others [35] have also found the Jungian classification important in relation to helping groups to think more effectively and innovatively about the futures they desire and how to reach them. Ozbekhan [36] has stated that redesigned systems rarely work well unless the controlling organisation itself changes, implying, in part, new values and beliefs.

There is a mood and a move towards a different sort of OR, more closely oriented to a recognition of behavioural issues. Decision-theorists, despite my earlier sceptical remarks, are by no means ignoring the difficulties that values and beliefs pose, not least the apparent inconsistencies that they give rise to; I refer in particular to studies by Lindley, Tversky and Brown [34], of reconciling probability assessments. It is not that I believe that there will be early or easy application of such theoretical approaches for the major social problems that OR is increasingly involved in, but the logical frameworks that stem from such work are bound to be of some relevance (apart from any direct applications of subjective probabilities to less complex problems in which values and beliefs are more tractable [2]).

Finally, I offer a tail-piece on fuzzy sets. There have been spirited interchanges, between Professor KD Tocher and Dr S French and his colleagues at the University of Manchester, in the Viewpoints columns of the Journal of Operational Research [43]. There is no need for me to defend Tocher's point of view, since he is doing very well on his own. However, there is one argument that French *et al.* use on which I must comment. "The problem with fuzzy sets", they say, "is that, to our knowledge, no-one has given an axiomatisation of the set membership function", and they ask for the assumptions on which this function is based. Here is the rub. If there were to be such an axiomatisation—if the rules for measurement were defined completely—there would be no vagueness. No doubt many would then be happy that conventional mathematical value theories could be followed, but they would no longer be treating fuzzy concepts. If fuzzy set theory is usable, it must allow for concepts (including membership functions) that are to be subjectively added to the logic for the purpose of a particular inquiry. The problem is not that one cannot do this, but whether unacceptable inconsistencies arise between the implications stemming from the application of fuzzy logic and the *intentions* of the transfer from vague verbal concepts to, say, membership functions.

By its nature, this paper itself has an essential vagueness. I cannot pretend to have understood fully the many references that I have read and cited. I do not know exactly what I am seeking, as a step toward an unattainable ideal, but I believe that I shall recognise hopeful landmarks when I see them. This paper aims to stimulate discussion in order to light up such landmarks, additional to the ones I refer to, and through debate to help create for others a more complete framework for the development of their ideas, whether or not these are in accord with mine.

The best we can do is to express our personal values on the subject of debate and hope that we can approach something which might be regarded as an organisational value, however fuzzily expressed.

Postscript

Since preparing the above, I have listened to the presentation and discussion of the paper by Lindley, Tversky and Brown [34] at the Royal Statistical Society. I remain unconvinced by the arguments. Tversky was not able to be present, which for me, was a pity, since he seemed to be the one most likely to have wished to stress the reservations made in the paper. The discussion raised interesting and important mathematical and philosophical points, but only one comment (that by Professor CAB Smith [41]) was strong enough to be consistent with my own reservations. Smith is not willing to accept a process that appears to him to treat people as if they were no more than instruments for the measurement of subjective probability: he thought that the authors were misguided, although he accepted the paper as being of theoretical interest.

I have now also read some unpublished papers by French [21–23] which extend the ideas of Lindley *et al.* French discusses, from the logical standpoint of a 'fully committed Bayesian', the interactions of values and beliefs that occur in group-decision making and in the transactions between decision-analysts and decision-makers. I find apparent inconsistencies in his philosophy, but nevertheless am greatly interested in his arguments and conclusions. In one paper [22], he suggests that a *consensus* should not be a distillation of experts' views presented as an agreed statement to a decision-maker, but that any consensus probability should be the decision-maker's own subjective probability, updated by him in the light of the experts' separate opinions and of his own beliefs in their reliability. This conclusion stems from French's proof that it is im-

Vol. 7
1979

possible to base a consensus rationally (i.e. in a Bayesian manner) on the information, beliefs and expertise within the group.

In the latest of the three papers [23], his statements about the things that motivate a group (such as common interests, "similar beliefs and aims", and a general concern for group stability) are at variance with my own experience. But, in any case, if people are to work as a group, they must experience the same problems of achieving 'rational' consensus as do the 'experts'. Further, any single decision-maker may not be rational in the Bayesian sense. It seems to me to follow that, if an analyst insists on his own values of rational analysis (as a Bayesian, or otherwise), he is likely to be an inadequate adviser to those of different persuasion (see also de Waele [44]).

In some ways, French is moving towards Eden's position [17, 18], although they are still far apart. The former's arguments all point to the severe difficulties of using normative theories to aid those who may neither understand nor accept these theories. In such cases, taking a behaviourist stance, the theories may be faulted because they are not found useful in practice. Further, French's own beliefs are bound to lead, at present, to arguments which depend on opinion, and he cannot, by his own thesis, expect the differences to be resolved rationally. Bayesian processes, as I have commented on elsewhere [10] are far from straightforward: the Reverend Bayes himself would not presumably be surprised that the beliefs implicit in them require an act of faith.

I have also had the benefit of discussing this paper with Professor White and his colleagues in the Department of Decision Theory, University of Manchester. I have left the text unchanged, although I know that I would now express some of my arguments differently and, I hope, more clearly; I also know that there is much more that I must study (for example, Keeney and Raiffa [30]) before I can make a thorough defence of my position against their decision-theoretic attack. On the other hand, if decision theory can be shown to deal adequately with values in practical decision environments the views I have expressed may urge theorists to demonstrate that this is so: they can best do this by themselves interpreting their mathematics in application, since third parties may well be distorting the logic. Meanwhile, the approaches to dealing with personal and organisational value systems that I wish to encourage can, quite independently, offer a way forward. This should not clash with theory: it should help both its development and its translation towards practice.

REFERENCES

1. ACKOFF RL & EMERY FE (1972) *On Purposeful Systems.* Tavistock, London, UK.
2. BALTHASAR HU, BOSCHI RAA & MENKE MM (1978) *Calling the long shots in R & D.* Paper presented at EURO meeting of EURO Working Group on Multicriteria Decision Making, York, UK, 2nd April.
3. BANNISTER D & FRANSELLA F (1971) *Inquiring Man,* Penguin, Harmondsworth, UK.
4. BANNISTER D & MAIR JMM (1968) *The Evaluation of Personal Constructs.* Academic Press, London, UK.
5. BENNETT PG (1977) Toward a theory of hypergames. *Omega* **5**(6), 749–751.
6. BENNETT PG & DANDO MR (1979) Complex strategic analysis: a hypergame study of the fall of France. *J. Opl Res. Soc.* **30**(1), 23–32.
7. BOWEN KC (1975) Helping with decisions on complicated systems. *IMPOS Bul.* V, 65–87 (Mexico).
8. BOWEN KC (1975) *Where techniques are lacking.* Paper presented to a Commonwealth Defence Science Organization Symposium, New Delhi, India (to be published in *IMPOS Bul.*).
9. BOWEN KC & SMITH DG (1976) Development of models for application to conflict problems. In *The Use of Models in the Social Sciences* (Ed. COLLINS L.), Tavistock, London, UK.
10. Bowen KC (1976) Some thoughts on Bayes' theorem, Bayes' postulate, randomness and fuzziness. *Jl Nav. Sci.* **2**(2), 145–151.
11. BOWEN KC (1977) *A systems approach.* Paper presented at a one day conference at the University of Aston, OR Society, Birmingham, UK.
12. BOWEN KC (1978) *Research Games.* ORSA text **3**, an OR Society Monograph with contributions by HARRIS JI, Taylor and Francis, London, UK.
13. BURTON JW (1969) *Conflict and Communication.* Macmillan, London, UK.
14. CLARKE D & RIVETT BHP (1978) A structural mapping approach to complex decision-making. *Jl Opl Res. Soc.* **29**(2), 113–128.
15. DANDO MR & SHARP RG (1977) Could OR be a science? *Omega* **5**(1), 89–92.
16. DANDO MR & SHARP RG (1978) Operational research in the UK in 1977: the causes and consequences of a myth? *Jl Opl Res. Soc.* **29**(10), 939–949.
17. EDEN CL (1977) Modelling the influence of decision-makers on the future. *Futures* **9**(4), 272–284.
18. EDEN CL (1978) Operational research and organization development. *Hum. Relat.* **31**(8), 657–674.
19. EMSHOFF JR (1968) Experience-generalised decision-making: the next generation of managerial models. *Interfaces* **8**(4), 40–48.
20. FREEMAN C (1969) Measurement of Output of Research and Experimental Development. *Statist. Reps & Stud.* **16**, UNESCO, Paris, France.
21. FRENCH S (1978) Updating of Belief in the Light of Someone Else's Opinion. Dept of Decision Theory, University of Manchester, UK.

22. FRENCH S (1978) Consensus of Opinion. *Notes in Decis. Theory*, **52**, Dept of Decision Theory, University of Manchester, UK.
23. FRENCH S (1978) Some Thoughts on Bayesian Decision Theory. Dept of Decision Theory, University of Manchester, UK.
24. GREEN RH (1978) Multidimension scaling for multi-objective policies: a comment. *Omega* **6**(3), 207.
25. HELMER O (1977) Problems in futures research: Delphi and causal cross-input analysis. *Futures* **9**(1), 17–31.
26. HELMER O (1966) *Social Technology*. Basic Books, New York, USA.
27. JAMES BAP (1975) Assessing force deployments: the decision maker's utility function. In *The Role and Effectiveness of Theories of Decision in Practice* (Eds WHITE DJ & BOWEN KC), Hodder & Stoughton, London, UK.
28. KAPLAN MF & SCHWARTZ S (Eds) (1975) *Human Judgement and Decision Processes*. Academic Press, London, UK.
29. KAPLAN MF & SCHWARTZ S (Eds) (1977) *Human Judgement and Decision Processes in Applied Settings*. Academic Press, London, UK.
30. KEENEY RL & RAIFFA H (1976) *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley, New York, USA.
31. KELLY GA (1955) *The Psychology of Personal Constructs*, Vols **I** and **II**. Norton, New York, USA.
32. LINDLEY DV (1971) *Making Decisions*. Wiley-Interscience, London, UK.
33. LINDLEY DV (1974) Discussion on the papers by Professor Tversky and Professor Suppes (Assessing Uncertainty; and The Measurement of Belief). *Jl R. Statist. Soc. B* **36**(2), 181.
34. LINDLEY DV, TVERSKY A & BROWN RV (1979) On the reconciliation of probability assessments. Technical Report TR77-10-25, Decisions and Designs Inc., McLean, Virginia, USA. *Jl R. Statist. Soc. A* **142**(2).
35. MITROFF IJ ET AL. (1977) The application of behavioural and philosophical technologies to strategic planning. *Mgmt Sci.* **24**(7), 44–58.
36. OZBEKHAN H (1976) *The future of Paris: a systems study in strategic urban planning*. Lectures to the OR Soc. and R. Soc., 28–29 October. (Also referred to in ACKOFF RL (1974) *Redesigning the Future*, P 31. John Wiley, London, UK).
37. RADFORD KJ (1977) *Complex Decision Problems: an Integrated Strategy for Resolution*. Reston, Virginia, USA.
38. RIVETT BHP (1977) Policy selection by structural mapping. *Proc. R. Soc. Lond.* A **354**, 407–423.
39. RIVETT BHP (1977) Multidimension Scaling for multi-objective policies. *Omega* **5**(4) 367–379.
40. SHAPIRO MJ & BONHAM GM (1973) Cognitive process and foreign policy decision-making. *Int. Stud. Q.* **17**(2), 147–174.
41. SMITH CAB (1979) Contribution to discussion on LINDLEY DV, TVERSKY A & BROWN RV. On the reconciliation of probability assessments. *Jl R. Statist. Soc. A* **142**(2).
42. SMITH DG (1978) On the role of operational research. *Omega* **6**(3), 208–209.
43. TOCHER KD & FRENCH S ET AL (1978) Viewpoints. *Jl Opl Res. Soc.* **29**(2), 179–182; **29**(11), 1132–1135.
44. DE WAELE M (1978) Managerial style and the design of decision aids. *Omega* **6**(1), 5–13.
45. WHITE DJ (1969) *Decision Theory*. Aldine, Chicago, USA.
46. WHITE DJ (1975) *Decision Methodology*. John Wiley, London, UK.

ADDRESS FOR CORRESPONDENCE: *KC Bowen, Esq. 37 Mountside, Guildford, Surrey, GU2 5JD, UK.*

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 513 to 519
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0513$02.00/0

Industrial Relations Problems—OR's Contribution[1]

JR MANN
GH MITCHELL

Operational Research Executive, National Coal Board, UK

(*Received April* 1979; *in revised form July*, 1979)

This paper considers the contribution which OR (Operational Research) can make to solving some of the problems in the industrial relations field. It first discusses the characteristics of industrial relations problems, categorising them according to the degree of conflict involved. The kinds of contribution OR can make are then discussed, first in a general way and then for each category in turn. The paper draws particular attention to the need for OR work in much of the industrial relations field to take account of the existence of conflict and of the game-like nature of the negotiating process. Some of the problems this presents to the OR worker are mentioned. Directions for future developments in the field are proposed. Finally there is a brief discussion about whether the issues raised in the paper are relevant to OR's contribution in fields other than industrial relations.

INTRODUCTION

INDUSTRIAL RELATIONS (IR) is the process by which privilege, power and wealth in an enterprise is shared out. Over time, the relative shares of these prized possessions move back and forth according to the actions of the parties concerned. The parties are traditionally management and trades unions but may easily extend to shareholders, banks, other companies, and, in the UK, national government. The actions these parties might use include strikes and other industrial sanctions, withdrawal of funds and so on. Many of these actions take place under a veil of secrecy with each party trying to move the frontiers of its influence in the process. To an outsider, the whole field is characterised by irrational, 'extremist,' even unacceptable actions by one or more of the parties concerned. Given this impression, it is not obvious that the field of IR is amenable to 'scientific method'.

Perhaps it is this impression, certainly emphasised in the press, which goes some way to explaining a limited interest in IR problems by the Operational Research (OR) practitioner in a company. Another reason may be the failure of some of the better known OR techniques to stand up to the problems encountered. However neither is a good enough reason for the field to be ignored. In terms of money alone, IR is undoubtedly a very important aspect of a company's activities and one where assistance ought to be valued.

In looking at the IR field, and the ability of OR to cope with the problems encountered, this paper is based on the viewpoint of the manager. However it is not meant to be 'one sided'. The ideas and methods explored could be equally useful for management in a Trades Union. In fact a fair amount of the paper considers how management in a company and in Trades Unions might be able to work together.

PROBLEMS FACING MANAGERS

We are basing our comments on management in an organisation with some level of Trades Unionism or similar grouping of employees. To some extent, we are also thinking about a fairly large organisation, that is one where management is in a sense removed

[1] A paper prepared for presentation and discussion at the Third International Research Conference on Operational Research and Management Science, organised by the University of Sussex and held at Bowness-on-Windermere, April 3–6, 1979.

from the 'shop floor' and where the number of employees is such that some element of planning is involved. However, the majority of problems mentioned are probably encountered in smaller organisations although perhaps to a lesser degree.

For the purposes of this paper, the problems facing management have been divided into three categories on the basis of the degree of conflict involved. This is obviously not a unique division, nor is it the only way that an OR practitioner might view the field. However we have the suspicion that it goes a long way to representing how a manager himself might see the problems in his work. The three categories are as follows:

Negotiative: These would be issues where the manager, as a representative of his organisation, is likely to be in open conflict with a Trades Union or other bodies. Typical examples are the setting of wage levels, length of working hours or manning levels. Common factors here are that the people involved have, or perceive that they have, conflicting opinions as to the best solution of the problem and therefore wish to have a voice in its resolution.

Semi-negotiative: These would be issues where, although there might be some element of conflict, there would be an overriding agreement by the bodies involved on some major objective. Typical examples would be the design of rosters or deployment systems where there is agreement among those involved to get the most convenient times/places of work. However there is unlikely to be full agreement on all issues, for example whether four long shifts are better than five short ones, and these could become a source of conflict. The common factors are therefore an overriding agreement on some main objective but enough potential sources of conflict for the Union, or other bodies, to wish to be involved.

Non-negotiative: These are issues where in the short term there is no conflict, or potential conflict, with other bodies. (Examples might be manpower planning for the future, diagnosis of why absenteeism is high or the maintenance of staff records.) Since they are not issues of conflict, the Trades Union or other bodies will not wish to be involved. (Very few issues indeed, of course, involve literally no conflict. One is tempted to think of issues in which the Trades Union has no interest as conflict-free, but over many such issues the manager is quite likely to find himself in conflict with other managers representing opposing interests within his own organisation.)

Any issue might, over time, change its status. Non-negotiative issues might become semi-negotiative if, for example, a Trades Union which had hitherto not challenged the way management kept staff records suddenly became concerned about some aspect of the information held on the record. A negotiative issue might, through the agreement of rules for settling disputes, become a non-negotiative one.

To characterise negotiative issues further, and to illustrate how they might present difficulties to the manager, consider some of the things with which such issues confront the manager.

High Stakes: The potential losses from a poor decision, such as conceding higher wages than a Union would accept, may well be very large.

Many participants: An issue such as wage levels may not be of interest to only two parties. It may be necessary to take account of the views of other Unions in the organisation whose differentials are under threat. Furthermore, outside bodies such as the government and major shareholders may wish to exercise some pressure in the negotiations.

Sectional Interests: It cannot be assumed that each of the participants will act as a united force. For example, various sections of a Union may have different views as to whether a certain pay deal is acceptable, thus, in effect, increasing the number of participants in the action. Additionally, the leaders of the Union may have certain personal interests such as endeavouring to get enough votes in the next election.

Incomplete knowledge: It is unlikely that our manager will be completely aware of all the objectives and possible strategies of the participants, in fact he may not know who all the participants are. His information will be partly based on fact and partly on his own perception of the views of the participants and may be seriously incomplete or even wrong in some circumstances.

A changing state of affairs: As the parties take part in negotiation, the manager will receive information that might alter his percep-

Vol. 7
1979

tions or add to his knowledge of the objectives and strategies of the other participants. This might suggest new strategies for himself. To act efficiently he will need a means of evaluating these strategies and so proceed, by iteration, towards an agreement.

Complexity: At any point in the negotiation, establishing the strategies of all parties concerned and assessing their implications is likely to be an extremely complex task.

Pressure: This is probably the aspect of the problem that affects the manager most of all, especially where there is much at stake. At one level, he will need a way of evaluating strategies quickly so as not to be overtaken by events. At another, it can affect the way he acts, forcing him towards a narrow, more stereotyped view of events.

It is interesting to compare the characteristics of the other two categories of problems—semi-negotiative and non-negotiative—with those given for negotiative problems above. In doing this it becomes clear that the categorisation is just a matter of degree. Where there is less conflict, there may be fewer participants or less pressure on the manager. In some cases incomplete knowledge may arise because of natural uncertainty. However there do not appear to be any hard and fast differences between the various categories. It is more a matter of how the negotiative aspects manifest themselves or need to be taken into account.

THE SUITABILITY OF OR

We cannot hope to give an exhaustive discussion and we shall concentrate on two aspects which we think are among the more important, namely the suitability of OR from the point of view of its approach to problems and of the way the OR practitioner works.

The OR approach

OR attempts to make a structured analysis of a certain state of affairs with the view to gaining some kind of improvement. Structured analysis necessarily involves some attempt to understand the state of affairs, the main issues and the mechanisms involved. It also involves some means of using that understanding to move towards an improved state of affairs. Whether improvement is achieved is of course a matter of the ability of the OR practitioner, the problem itself or luck.

The OR approach involves model building because building a model is the means of acquiring and setting down a structured understanding of the state of affairs. In some cases, the model may incorporate a device for producing answers to a problem on the basis of the understanding embodied in it. If this device is to be useful the model incorporating it must embody all the important facts of the process within which the problem is set. There is no reason in principle why the model should not be fully adequate. However human model builders and computers have practical constraints. This inevitably leads to a degree of simplification in the model. Simplification might mean the model fails to deal with conflicting objectives, to recognise all of the participants and the various sectional interests and to deal with the behavioural side of the problem. The model, viewed solely as a means of acquiring structured understanding, is a more flexible concept. It is a means of structuring our attempts to understand reality, our possible actions and their effects. Such a model serves to stimulate and guide our thinking of new actions, point out areas of ignorance and thus guide our research. There is a large element of this in the device type model, but here it has a value in its own right.

Both kinds of model have certain characteristics which can affect their suitability to the problems in the IR field. Models often incorporate, usually implicitly, beliefs about criteria to define the best actions to be taken, for example the maximisation of company profits. This type of model is only useful where there is some consensus amongst the participants that the action is the best one, an unlikely state of affairs where there is a high degree of conflict. To be of greatest use, the model should be value-free in the sense of being as free as possible of those values which differentiate the participants. The model will also need to be flexible and not dependent on the position at the time it was built. As new strategies and interests manifest themselves, the model must be able to cover them. This will affect the ability of the model to produce information quickly and to help the manager acting under pressure.

These requirements of models if they are to

be usable in the IR field are at least challenging and may well be impossible to attain in the short run. In our experience there is not always the prospect of quick high returns from model building and a patient sponsor might be an essential requirement.

The role of the OR practitioner

The OR practitioner can adopt one of the following positions:

- He can work for one side. This is the traditional role of the OR man who works for management in an organisation.
- He can work for both sides. This is possible on the semi-negotiative issues where there is some consensus of objectives, but infeasible in states of high conflict.
- He can work as an observer. In this case, the OR practitioner would probably not be employed by one of the participants. His aim would be to learn about the process, possibly in the hope of influencing one or more of the participants in the long run.

Thus, unless the problem is considered to be suitable for combined study by the participants involved, the OR practitioner can only work for one of the participants or in the outside role of an observer. However as an observer he cannot expect to have any direct influence over the resolution of the problem in question. The point here is that the OR worker will, to play an effective role, probably have to compromise about some of his ideals as an OR worker. In particular he might be faced with the choice of giving up the ideal of objectivity (by working for one side) or of purposeful intervention (by settling for the observer role).

STATE OF THE ART

Negotiative problems

Let us first look at the negotiative problems such as wages bargaining. The resolution of this type of problem has three phases: preparation, negotiation and termination—the first two at least are, of course, interwoven. The preparation stage is not a matter of conflict, unless there is conflict within the manager's own organisation. If there is, then the preparation becomes a negotiative issue in its own right. Otherwise it is non-negotiative and will be dealt with below. The third stage, termination, is likely to be semi-negotiative with all parties wishing to work towards a common goal.

The negotiation itself is a formal exchange of information as a result of which new strategies arise and the manager becomes more aware of the options and preferences of the other participants. In order to help the manager, the OR practitioner needs a means of using that information for the benefit of the manager for whom he is working. To do this, the practitioner will need a versatile, preferably relatively value-free and fast-acting model. Moreover it must recognise and be able to deal with the conflict between the participants.

The models which satisfy these requirements, hypergames and metagames, treat the problem as a game. Both types of model give the manager a structured understanding of the issues by comparing his own strategies with those of the other participants. Game theory alone is not suitable since it assumes that each player is well informed of the structure of the game, that is he knows the strategies and preferences of the other players. Hypergame theory [1], however, allows for the possibility that our manager may only perceive a small proportion of the permissible strategies, may not be aware of the preferences and may not even known who all the participants are. The importance of the theory is simply that it reminds us that other people might be playing a different game from the one we think we and they are playing. Emphasis is placed therefore on trying to understand the perceptions of the other players—an understanding that one hopes will increase as the game is played—so as to think up and deploy better strategies than would arise from consideration of our own possibly narrow perception of the game. Unfortunately a complete hypergame formulation of a normal bargaining process can be very unwieldy given all the options and combinations of options. The OR practitioner will therefore need to work closely with the manager, presenting the information to him in a form easy to assimilate.

Options analysis [4, 7, 8], an extension of metagame theory, is based on similar ideas. It too has been used in an organisation undergoing an industrial relations dispute. To use this technique it is necessary to have a list of all the interested parties and, for each, a list of the

Vol. 7
1979

actions, or options, available to them. Each option is taken or not. The method seeks a combination of options that the player should take, his strategy, in order to secure stability —a position from which no player can move to advantage. Of course, at the outset options for all the players cannot all be listed, and the analysis might well change shape as a fuller range of options is revealed through time. The preferred strategy might also change as another player's strategy is partly revealed through an action he takes. Again, the analysis can become somewhat unwieldy but the approach is one that the manager can understand and which probably conforms to his own way of looking at the problem.

A major question about gaming methods and conflict analysis is whether their use is of itself helpful. Unquestionably they have an intuitive appeal to anyone who has been involved in negotiations, if only because of the powerful abstract description they offer of a process with which the negotiator is familiar but has never seen modelled before. However their effective use calls for creativity in developing new strategies and options and this creativity and acceptance of its output might be much more a consequence of the personal skills and attitudes of the analyst and manager than of the use of the methods. It is quite possible that the methods will simply underwrite and rationalise stances that would be adopted anyway: the good negotiator may not need the methods and the bad might not be able to use them. There seems to be, more than in most kinds of OR, an act of faith in advancing the use of these methods.

We are not aware of any other types of models which fully appreciate the complexities which must be faced in negotiative problems. Of course it is possible to build useful models to predict the cost and implications of the various strategies used, but these models alone will not advise the manager on how to use those strategies in the negotiation. They are merely for preparation.

Thinking about the observer, his role is to study one or more of the participants, analyse their behaviour and test various hypotheses probably with the aim of enlightenment. HG Jones covers this field in some detail in a recent article in JORS [5]. The techniques involved would be to a large extent those used in behavioural research and economics. There is also scope for sociological and anthropological research. However it is not obvious that any of the information so gained will be of immediate interest to the manager.

Semi-negotiative problems

The common factor of semi-negotiative problems is the existence of some common ground which allows for joint examination of the state of affairs. Once this common ground is established and exploited, the actual means of problem solving is not an issue for this paper. In fact the problem, when it is defined, may not be an 'IR problem' but a question of scheduling, production line design and so on. The IR element is the avoidance of conflict and the establishment of a situation in which the parties can work together and achieve implementation.

A way of doing this is through a procedure known as action research [3]. This method was first used sometime before the war and is often used by behavioural scientists. The method is to set up a task group consisting of people doing the job around which the problem centres, their supervisors, trades union representatives and any people who might be immediately affected by the outcome. This group could be set up by the OR practitioner in conjunction with management. The group would then discuss the problem, providing information for the OR practitioner in his subsequent analysis. The group has a say in the information collected and the use of results. In fact, by virtue of this process, implementation tends to be much easier to achieve.

This technique can often succeed where others have failed. For instance, in the NCB a statistical study of absence proved to be of limited use. However it transpired that the union shared with management the desire to reduce absence. Action research was therefore thought a suitable line of attack. Its use pinpointed the causes of absenteeism and thus went some way towards alleviating the problem. The method has been used widely by Henley [2] at companies such as United Biscuits. A similar example is the rostering study at British Airways [9].

Non-negotiative problems

In this category fall all of those problems which can be tackled wholly within the com-

pany. Currently, in most cases, it would include manpower planning, staff data banks and design of recruitment and training procedures. The field is too wide to explore in detail. The techniques developed for dealing with these issues have certain drawbacks but on the whole these are due not to any aspect of IR but to the problems of handling uncertainty and technological change, of how best to categorise staff and of handling large amounts of information.

The important aspect of the problems under this heading is that they do not involve conflict at the time they are considered. The company alone can decide on the objective, the assumptions to be made and the degree of simplification in the models used. However it follows that the use of these models should be questioned as the position changes, their limitations being recognised.

If participation increases, it is questionable how much some of the subjects mentioned will remain non-negotiative. For instance, there may be a need to make all or some part of a staff records system accessible to the workforce. Canadian Airways [6] have done this. However, we think it likely that all parties will continue to have private models as part of the planning process.

THE FUTURE

The developments we would like to see are:

- More joint problem solving where the ground exists. This could be achieved by using action research or similar methods.
- More use and further development of models whose aim is to give a structured understanding of the conflict process, for example gaming models.

To some extent we would wish to encourage the role of the observer. However, from the point of view of the manager, this is not likely to help him as much as the two developments listed above. The development we would not wish to encourage is the construction of models which simplify the process, ignore the behavioural aspects and do not, where necessary, recognise conflict.

To achieve and use these developments, there is a need to achieve:

- A better understanding by OR practitioners of the Industrial Relations field, the participants and their culture. This could be achieved through OR education and through a determined attempt to work in this sphere.
- Close contact with management. IR work can involve a certain aspect of secrecy thus requiring a fair amount of trust between manager and OR worker.
- A winning over of management. Managers themselves may not see IR as a suitable field for the OR practitioner because they see it as being 'unscientific'.

The third point is particularly important. We think that the confidence of management may be most easily achieved if OR can present methods and techniques which the manager can trust and understand.

IS IR UNIQUE?

Before we conclude, it might be interesting to compare IR problems with other management problems, in particular those which attract most attention from OR practitioners. It is tempting to think that IR problems differ from others in an organisation but in fact any differences may only be slight. As with the problems inside the IR field, it is a matter of the degree of conflict with other bodies and the degree of interest they have in the resolution of the problem. Many issues in industry are of the negotiative type. Indeed we have the sense of negotiation pervading much of the work of the OR group to which we belong.

Consider for example the problem of choosing a large item of machinery. The typical approach to this problem would be to compare all machines which suit the requirements of the job in terms of capital costs, running costs, reliability, ease of operation, manpower savings and so on. The decision on which machine to advise management to buy would be based on some kind of DCF (Discounted Cash Flow), cost-benefit or trade-off analysis. The manager may decide to act on this advice but we know in a number of cases that he might not. This probably is not because the manager is at fault, more often it is because he sees the issue in a wider context. For instance, he may wish to avoid a certain supplier getting a monopoly of the market, he may wish to avoid conflict with the unions over manning levels or he may wish

Vol. 7
1979

to buy more prestigious machinery than his competitors have. To a great extent this is saying the manager has a more complicated objective function than was first thought, indeed perhaps a multi-dimensional one. However there is quite an element of the manager seeing the decision as a statement in some continuous negotiation going on with various parties, a statement which he believes or fears might draw responses, some of which would be unfavourable to him. These aspects of the decision would be difficult to build into the kind of model we have mentioned above.

It is not so much a case of failing to define the right problem; more it is one of failing to take some of the more fundamental issues into account. In a case such as this, an approach like options analysis may have produced results more in keeping with the state of affairs as perceived by the manager.

This example may not be unusual. It leads us to suspect that some of the methods appropriate to the IR field may be particularly useful in other parts of an organisation. Moreover, for complex problems with some degree of conflict, the IR field may be a very good testing ground.

CONCLUDING REMARKS

This paper has outlined the problems in the IR field in terms of the degree of conflict and involvement of participants both within and outside an organisation.

It is concluded that the preferable approaches are those that aim to give a structured understanding of the conflict and the strategies of the participants, thus allowing a manager to consider thoroughly his position in the process. It is also argued that IR problems may not be significantly different from those in other branches in an organisation. If this is so, methods developed for tackling IR problems may prove to be of great use elsewhere. Furthermore the IR field may be a useful testing ground for these methods since it is, in some sense, the extreme case.

(The views expressed in this paper are the authors' and not necessarily those of their employer.)

REFERENCES

1. BENNETT PG (1977) Towards a theory of hypergames. *Omega*, **5**(6), 749–751.
2. BIRCHALL DW, CARNALL CA & WILD R (1976) *The development of group working in biscuit manufacture: an evaluation of outcomes.* Paper presented to the Workshop 'Humanisation of Work', EISAM, Brussels, Belgium.
3. CLARK AW (1976) *Experimenting with Organisational Life.* Tavistock Institute, Plenum Press, New York, U.S.A.
4. HOWARD N (1975) The analysis of options in business problems. *Infor* **13**(1), 48–67.
5. JONES HG (1978) Further areas for OR modelling in industry. *Jl Opl Res. Soc.*, **29**(3), 191–198.
6. LITTLE D (1977) *Human resource planning in Air Canada.* Paper presented to Manpower Soc. Conf.
7. RADFORD KJ (1975) *Managerial Decision Making.* Reston, Reston, Virginia, USA.
8. RADFORD KJ (1977) *Complex Decision Problems: an Integrated Strategy for Resolution.* Reston, Reston, Virginia, USA.
9. WILLIAMS M (1977) *Participation for change or entrenchment of the status quo?* Paper presented at OR Soc. Conf.

ADDRESS FOR CORRESPONDENCE: *Mrs JR Mann, Operational Research Executive, Coal House, Lyon Road, Harrow, Middlesex HA1 2EX, UK.*

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci.. Vol. 7. No. 6. pp. 521 to 526
© Pergamon Press Ltd 1979. Printed in Great Britain

0305-0483/79/1101-0521$02.00/0

Increasing Effectiveness of Modeling

MYLES A VOGEL

Becton-Dickinson & Co., New Jersey

(*Received March* 1979)

Modeling is the essence of the operations research study. Yet with the exception of rare instances does the model of the study realize its full potential benefits in the business world. In this treatise an attempt is made to familiarize the business world with the types of models that exist including their operational and functional attributes. In order to increase the model and or study effectiveness the essential elements in performance are expounded to illustrate the comprehensiveness required. In conclusion, the recognition that modeling is not a utopia and that many pitfalls exist is explored with recommendations made to diminish or eliminate the negative impact of these shortcomings.

THE TITLE of this paper in essence equates to saying 'increasing the effectiveness of operations research.' Operations Research or Management Science whatever you want to call it has not realistically lived up to its initial expectations. Originally looked on as the savior of management, it has been anything but that. Formal programs have been developed at most large universities so that a student can get a bachelors, masters or doctorate degree in Operations Research, Business Science, Management Science, etc. Universities are cranking out MBA as an indication of status. At one time quality and knowledge prevailed, but because of economic, social and governmental pressures quantity presides. This assembly line mentality could foster the downfall of Operations Research if continued.

It is quite impressive for an MBA student to be able to set up a long involved LP program or a computer model that only the 99th percentile of the Operations Research technicians could understand.

But I must ask if the model is not implemented and utilized at least to stimulate management thinking or the management process, what value do we get?

I am not deploring research, but whether it is new product research, computer research, etc., in business the bottom line is what counts. The earnings per share (EPS) is the key number in many businesses. Businesses/Corporations are extremely sensitive to this ratio. EPS stimulates positive reports from Wall Street. Steady increases in EPS assist in obtaining funds from institutions investing at a somewhat lower interest rate! A company is looked on as minimal risk. This also serves a competitive advantage in the marketplace.

In essence the businessman is very sensitive to the bottom line. If Operations Research is to improve its position it must impact positively the bottom line and in turn react in a real time mode.

Operations Research practitioners should start off with short term projects, fit the project/technique to management needs, do not stipulate a technique and then apply the application. The key is for the Operations Research analyst to build credibility with the user. Build confidence by reacting in real time, with concise information and knowledge of the operation. Even a small addition error can negatively impact user confidence in the analyst and his interpretation of the project.

I do not want to project a cloud of doom. Certain groups have performed extremely sophisticated types of analysis on projects such as negative income tax, skylab projects, etc. But this is a small minority. Look into your

operation and your implementation ratio, review the time lag to get back to management with the answer.

Keep a log indicating where this bottleneck occurs. Usually a great deal of the problem resides in the electronic data processing (EDP) organization or with your immediate supervisor being apprehensive about submitting your conclusions. Operations Research has a tendency to pry, delve, and uncover things in an operation that management might wish to leave alone. One must use discretion in this case. It also appears that the closer the Operations Research group is near the top the greater the probability that implementation will occur. Even one or two levels to work through can lead to significant success, but when we get to the third, fourth, and fifth levels the analysis gets lost along with project definition, etc. Usually, these projects are used by the department manager as 'in case' analysis—'In case the president should ask.' My advice to the analyst after 2–3 years in this sort of environment is to leave. What occurs is the repetition of the same projects utilizing the same techniques and you are tunneling your experience, not learning new knowledge.

One individual I recently talked to had 10 years' experience in Operations Research at the same level. He really did not have 10 years' experience but 2 years' experience five times. This is not to say 10 years' experience in Operations Research is of no value, if the individual is growing in responsibility or exploring significantly new areas, then 'OK'. But this is not generally the case in large Operations Research departments and stagnation can set in.

But let's get away from this doom and gloom scenario and investigate ...

What is an Operations Research project?

What are its characteristics?

There are a myriad of ways to approach a management problem; many of these are related or at least somewhat associated. In my mind there is certainly no dogma that indicates that linear programming techniques should be used for problem type *A* or queueing theory always solves problem type *B*.

Just as there are many Operations Research techniques, many types of specialists such as Operations Research Analysts, Industrial Engineers, Economists, Accountants or Financial Analysts all can approach a management problem in various ways. There is no clearly established guidelines for technique application or discipline that specifically is the best manner to derive a solution.

But most Operations Research applications do possess some distinguishing characteristics and I offer four as a frame of reference.

OPERATIONS RESEARCH APPLICATION ATTRIBUTES

1. *A primary focus on the decision-making*

Principal results of the analysis must have direct and unambiguous implication for executive action. The reason we go through an Operations Research study is to lead to or assist in the decision making process.

2. *Analysis based on economic effectiveness = optimization, maximization or minimization*

The Operations Research approach revolves around development of a myriad of feasible solutions based on measurable—quantitative values. The solution is normally directed at maximizing, minimizing or optimizing some measurable quantity to the business firm. These quantities may be net income, cash flow, return on investment (ROI) etc.

3. *Reliance on a model*

Normally this model is mathematical. But let me caution it should not, if at all possible, be too complex if the hope for understanding and implementation prevails.

Model building is the essence of the Operations Research study. Constructing a model helps put the complexities and possible uncertainties attending a decision making problem into a logical framework amenable to comprehensive analysis.

Such a model classifies the decision alternatives and their anticipated effects, indicates the data that are relevant for analyzing the alternatives, and lead to informative conclusions. In short, the model is a vehicle for arriving at a well-structured view of validity.

4. *Utilization and dependence on the electronic computer*

The Operations Research model, because of the requirement for analyzing alternatives, sometimes called 'What If? Scenarios,' depends quite heavily on the electronic computer. The

Vol. 7
1979

computer is an aid to the Operations Research analyst so that he can analyze the complexities of the mathematic model, manipulate data, or for computation.

I believe the main characteristics of an Operations Study are now distinguishable but what approach should be taken? If the Operations Research study is to be successful, a systematic approach and user interaction must take place.

The analyst and the user (decision making) must get together and ascertain a clear purpose for the analysis or model; with the end use being a guide for the building and development of the model.

The outline listed below is a guideline to aid in a systematic approach to modeling or the Operations Research study. [7]

SYSTEM DEFINITION

A modeler must start with a clear definition of the questions which the model is being designed to answer. This top-down approach can be used to uncover the components of the system which need to be included in the model. After defining these components, the modeller can define the parameters and variables describing the way these components interact and the functional relationship between them. This definition of architecture is often the most difficult part of the process: it is critically dependent on the purpose the model/study is intended to serve.

MODEL FORMULATION

The next step in refining the model is to define in a formal way the interrelationship among the components of the model. Remember the decision maker is keenly aware of these relationships and must be the key input in this phase. By using simple flow charts, block diagrams etc. you can put into layman's terms the intricacies of the model. These first two phases are the key and require considerable user/analyst interface.

DATA PREPARATION

In most situations the data of interest do not exist or at least do not exist in the form required. Make sure you get the data you require. Do not make assumptions just to get the data in order. Take time and get the data you need.

MODEL TRANSLATION

Once the basic architecture has been built and the correct data have been prepared, it is then a question of translating the model into a form which can be computerized. (Although this is the first time a computer enters the picture in a direct way, quite often a lot of us start with this step). Translating the model for the computer involves the use of a computer programming language. There are now literally hundreds of programming languages in use, and the choice of the most appropriate language is often a nontrivial decision. Many times, however, the language is chosen and the model translation begun before the model has been defined.

VALIDATION

After the model has been translated and installed in the computer, it must be validated—a nontrivial task. A good vantage point for validating the model is to think of it in terms of the model-builders view of how the system works and to consider the question of truth as irrelevant. Validation tests are concerned only with the model's ability to answer the questions poised.

STRATEGIC PLANNING

Since a computer model costs money to run, it is necessary to plan the runs on the model very carefully. This strategic planning, of which trials are going to be run should be done very early in the model-building cycle.

TACTICAL PLANNING

Detailed plans must be made for running the model, including considerations of starting conditions, how long to run the model for equilibrium conditions, number of runs, etc.

INTERPRETATION

The answer that comes out of the model must be understood by the user; must give reasonable answers for reasonable situations;

and must give implementable solutions. The model must have realistic data requirements, must be able to answer 'What if?' kinds of questions, must be easy to modify, and must be cost effective. In short, it has to be a saleable product.

DOCUMENTATION

Not only must a program work, it must also be described and documented so that another analyst can easily understand and take up the slack.

In summary, only with a disciplined approach as indicated will the process of model building/Operations Research study change from an art to a science and so be more acceptable to the community at large.

But I am sure you are all sitting there stating that you know the characteristics of an Operations Research study and you follow a systematic approach to model building. Yet implementation at large is still negligible to the whole.

There are still several factors besides not knowing or acknowledging the Operations Research study characteristics or performing the analysis in a systematic manner that have impacted negatively on the Operations Research effectiveness ratio.

These negative elements are:

(1) Shortage of Time.

(2) Inaccessibility of Data.

(3) Resistance to Change.

(4) Long Response Time.

(5) Invalidating Specifications.

These factors occur in many forms: [3]

1. *Shortage of time*

Management scientists simply do not sufficiently understand the constraint of time on decision making, and particularly on decisions that count; and the techniques they develop reflect that fact. They may write about time as a limitation. They may admonish managers for letting time push them into a 'crisis' mode. They may recognize the constraint of time with a few words and comments on its influence. They may say that they, too, experience time constraints. But their techniques are so time consuming to use that managers pass them by.

This simply means that the management scientists (a) need to get out of their relatively unpressured worlds and experience the impact of time on the decision making process, and (b) need to build the time factor into models instead of leaving it as an exogenous variable.

2. *Inaccessibility of data*

The second reason for ignoring management science in practice is relative to the time problem. A manager will ordinarily use data or a management science tool only if both are conveniently, speedily accessible.

What does the management scientist do? He continues to construct models that call for substantial investment in design and data collection and managers discarding them. The statement is made that most data are not in the form that most models call for, or that they are not complete, yet the management scientist goes right on calling for inaccessible, non-existent, or uncomputed data to suit 'theoretically correct' models. And hence managers continue to say, 'skip it.'

Why shouldn't the management scientist design a realistic model, or a realistic part of a model, or come up with a realistic data prescription? The result might be extremely crude, it might be embarrassing to the theoretician, it might be shot down by the purist and the theoretician. But it just might be used.

3. *Resistance to change*

Educating others in the organization who are not familiar with the tools, and who would resist using them if they were, is just too difficult a task.

Management scientists typically regard this problem as outside the scope of their jobs, at most, they remark on the need to educate more people and to change organizations so they become more scientific.

Realistically the management scientists should get themselves some education and direct experience in the powers, politics and change-resistance factors in the real world of management so they can better incorporate the imperfect human variables in their work.

Vol. 7
1979

4. *Long response time*

Fourth, few management scientist people are geared up to respond to significant management problems in 'real time.' Most management scientist personnel in staff positions of live organizations operate on a time frame that is slower than that of the line manager. It is their nature to approach a problem in a methodical, thorough way, even when the required response time dictates that they come up with a 'quick and dirty' solution.

5. *Invalidating specifications*

Fifth and finally, it is standard operating procedure for most management science people to strip away so much of a real problem with 'simplifying assumption' that the remaining carcass of the problem and its attendant solution bear little resemblance to the reality with which the manager must deal.

The time constraints, the data-availability questions, the people problems, the power structures and the political pressures are the important, nasty areas that close to the essence of management are simplified out of existence so that a technically beautiful and useless, resolution may be achieved.

They leave it to the manager to worry about shortage of time, inaccessibility of data, resistance to change, slow response time and oversimplified solutions.

But the ineffectiveness of Operations Research is not totally in the corner or domain of the Management Scientist. Even when *all* the previously mentioned factors are overcome, the manager offers an almost complete contrast. He integrates and assigns the sizes, he sees situations as mosaics, his thoughts and decision processes are so complex you can never be sure how he is operating.

At the core of his value system are depth and breadth of experience which he may even permit to outweigh facts where the whole picture seems to justify it.

But how can we overcome this gap which is not only cultural but practical and which affects Operations Research effectiveness?

Closing any gap requires that one or both cultures change. It is my belief that the management scientist must move first. The end product is supposed to be management, after all, not management science.

The management scientist is not going to find it easy or comfortable to change and yet it is he who must change most in attitude, action and life style. He is going to have to think in terms of managers'

—perception,
—needs,
—expectations, and
—pressures,

that is, if he wants to have impact on the real world of the manager.

What concretely can be done? Let me offer a few suggestions to the management science people and the managers they are supposed to be helping.

First, top management should not isolate the management science people but sprinkle them throughout the organization in situations where they can really go to work. It should give them line responsibility for results. Management should not think of having an operating person on a management science team—it should think of having a management scientist on an operating team.

Second, managers should demand implementation by management scientists; they should not tolerate 'package' solutions that leave out the complicating factors. In that way, managers can avoid simplistic, unworkable solutions that ignore the factors of time, data accessibility, unwillingness of people to change, power and so on.

Third, even when professional management scientists are brought into companies as consultants, they are often given the easy, old problems, for the reasons that I have named. This expectational cycle has to be broken.

AT THE UNIVERSITY

The same general approach is valid within universities.

First, both management science faculty and students have to get out of the isolated, insulated world of academe. They must go beyond structured cases and lectures and become directly involved in real-world, real-time, live projects, not as a way of applying what they know, but as a way of learning.

It is a mistake to teach the student linear programming or decision theory and then search for a problem to which the tool can be

applied. That creates the classic academic situation from which managers are revolting—the tool in search of a problem. Instead, tackle the real problem. This will be frustrating, but the frustration of trying to reach a workable solution can be used to teach the management scientist or student in a way that is useful both to him and to the business or government unit. The solutions thus derived may not be so elegant as they might be, but they may be used. The student who wants to reach for higher, more sophisticated theories should be treated as a special case.

Second, management science people should stop tackling the neat, simple problems or refining approaches to problems already solved. These projects may be easier, but working and reworking them will not help bridge the cultural gap I am talking about.

Instead, tackle the tough problem. The management of time pressure and the use of the persuasion and negotiation required by a real, tough problem will give both the faculty member and the student some discipline in convincing others to follow a strange idea, to cooperate and to listen.

Finally, universities should bring in real managers and involve them directly in problem-solving and joint-learning sessions. The universities can offer managers an update in science; corporate managers can offer universities an update in management.

These are some of the ways to build bridges. There are other ways to tear them down or to maintain the gap. Jargon, for example, will drive away managers. So will intellectual snobbery toward 'intuitive' decision making. Management scientists should dispense with both. Managers can maintain the gap by continuing to refer past disillusionments and never allowing management science to profit from past mistakes.

REFERENCES

1. ACKOFF RL (1960) Unsuccessful case studies and why. *Ops Res.* **8**(2), 259–263.
2. GINZBERG MJ (1978) Steps towards more effective implementation of MS & MIS. *Interfaces* **8**(3), 57–63.
3. GRAYSON CJ (1973) Management science. *Harvard Busin. Rev.* **51**, 41–48.
4. HALL JR. JR (1978) OR/MS: dead or dying? RX for survival. *Interfaces* **8**(3), 42–52.
5. HOWITT WS (1978) Value of the simple approach. *Interfaces* **8**(3), 64–70.
6. KOLB DA & FROHMAN AL (1970) An organization development approach to consulting. *Sloan Mgmt Rev.* **12**(1), 51–65.
7. SHANNAN RE (1970) Simulation, modeling and methodology. *Simulation* **IX** *Newsletter*, November.
8. ZARD DE & SORENSEN RE (1975) Theory of change and the effective use of management science. *Admin. Sci. Q.* **20**(4), 532–545.

ADDRESS FOR CORRESPONDENCE: *Dr Myles A Vogel, Manager; Operations Planning & Analysis, Becton–Dickinson, Rutherford, New Jersey 07070, USA.*

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 527 to 537
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0527$02.00/0

Improving MIS Project Selection[1]

MICHAEL J GINZBERG

Graduate School of Business, Columbia University

(*Received January 1979; in revised form April 1979*)

Organizations develop and install computer-based information systems because they expect their benefits to exceed their costs. Benefits come in all varieties, but our measurement skills are best developed in the area of efficiency improvement—reducing the cost of data processing. Since there are normally more potential projects than available resources, this imbalance in our capacity to measure the full range of benefits can lead to accepting projects of lower overall value than that of some which are rejected. This paper analyzes the impact of present benefits measurement practices, and suggests some possible improvements.

BENEFITS MEASUREMENT: THE BASIC PROBLEM IN PROJECT SELECTION

GAUGING the benefits of management information system (MIS) projects is not a new problem. It has been discussed for many years; yet it remains, for the most part, an unsolved problem. The usual focus in discussions of MIS benefits is on the evaluation process, measuring benefits after the system has been implemented. This paper focuses on another phase of the MIS development process—project selection. The criteria used to evaluate a system after implementation should be the same as those used to select that system for development in the first place. Thus, selection and evaluation are two sides of the same coin. And in many respects, selection is a more difficult problem.

Early efforts to use computers to support the management of organizations focused on automating clerical tasks. The primary benefit of these systems was a reduction in the cost of data processing. The use of computers has now moved well beyond clerical automation, and this has greatly expanded the range of benefits. Knutsen & Nolan [7] suggest the following six classes of benefits arising from computer-based systems:

(1) equipment displacement,

(2) reduction of personnel in data processing tasks,

(3) increased operational efficiency in functional areas,

(4) increased sales,

(5) better managerial planning and control, and

(6) other organizational impacts—e.g. flexibility.

Every system is likely to offer a unique mix of these six types of benefits, and this is the source of the measurement problem. We must find techniques which allow the comparison of systems offering essentially different types of benefits.

Some measurement of system benefits is required at both the selection and evaluation stages, but the nature of the necessary measurement differs due to differences in the decisions being made. Four distinct decisions can be identified:

(1) go/no go (investment) decisions,

[1] This research was supported in part by a grant from the Faculty Research Fund of the Graduate School of Business, Columbia University.

TABLE 1. DECISIONS REQUIRING SYSTEM EVALUATION

Decision	Implicit question	Measurement demands
Selection decisions		
1. Go/No Go decision	does this project meet the minimum requirements for consideration?	identification of benefits sufficient to meet minimum required level
2. Resource allocation decision	does this project represent the "best" use of resources among the available (and potential) uses?	identification of *all* benefits; measurement on a scale which allows comparison across projects
Evaluation decisions		
3. Post project audit	is this project providing the promised benefits?	tracking of all *expected* benefits; measurements need not be comparable
4. System tuning	how could this system be modified to provide greater benefit to the organization?	identification of *all* benefits; measurements need not be comparable

(2) resource allocation decisions,

(3) post project audits, and

(4) system tuning (performance diagnosis and improvement).

The first two of these decisions relate to project selection, and the last two to after the fact evaluation. Table 1 outlines these decisions, the implicit questions upon which they are based, and the demands they place on the measurement of benefits.

Measuring system benefits becomes more difficult either (1) when it is necessary to measure *all* benefits or (2) when benefits must be measured on comparable scales. For the go/no go decision on a single project, neither of these requirements apply. For after the fact evaluation decisions it is often necessary to consider the full range of benefits, but comparable scales are seldom required.

It is only in the case of project selection in a limited resource environment—the Resource Allocation Decision—that it becomes critical to measure all types of benefits *and* to use comparable measures for all types. If only a few organizations needed to face this question of resource allocation, the need for improved project selection techniques would not be great. However, this is not the case. Most organizations have more opportunities than available resources [12]. Thus, the value of improved techniques for project selection could be substantial.

A NORMATIVE FRAMEWORK FOR PROJECT SELECTION

Ideally, the solution to the project selection problem should be the application of cost-benefit analysis. Carlson [3], in reviewing six selection and evaluation techniques, concludes that cost–benefit analyses should provide more meaningful results than any of the other available techniques. He notes, however, that there are often problems in obtaining measures of project benefits.

The existence of these measurement problems led Knutsen & Nolan [7] to reject cost–benefit analysis as a valid tool for MIS project selection. They contend that the technique's focus on monetary measures of benefit will lead organizations to choose projects which are clerical in nature, because the resulting cost savings are easily measured. Thus, they argue, projects in areas of great importance to the organization will not be undertaken because their benefits are more difficult to measure.

Knutsen & Nolan's point should not be ignored. If the use of cost–benefit analysis results in ignoring the harder to measure benefits, it will often lead to unwise decisions. We must ask, however, whether this is a necessary result of using the technique.

King & Schrems [6], in a tutorial on cost–benefit analysis for MIS, suggest that the technique can be used to consider a wide range of benefit types. They point to certain measurement problems, but they do not believe these problems are insurmountable. Indeed, they

Vol. 7
1979

describe the ways in which the dollar value of certain hard to quantify benefits has been estimated in past projects.

To be sure, there are difficulties in estimating the value of some project benefits. King & Schrems [6] list a number of these problems, including)[2]

(1) the 'natural' unit of measurement may not be comparable across all benefits;

(2) some benefits will be of different value to different people (e.g. different users of a multi-user system);

(3) the quantification of some benefits is, of necessity, highly subjective and subject to great uncertainty;

(4) the benefits actually obtained may depend on the operating environment for the system (e.g. response time of the system may affect the value of the information it produces); and

(5) benefits are estimated at the start of the project but may change over the life of the system.

Their implicit response is that while these problems make the proper use of cost–benefit analysis difficult, they do not make it impossible.

Indeed, if we consider the types of benefits claimed for MIS projects (Knutsen & Nolan's list, for example), it *should* be possible to quantify all of them. Each of these benefits is aimed at increasing the efficiency or effectiveness of the organization. Ultimately, such changes are expected to have a financial impact on the organization. In some cases—e.g., equipment displacement, personnel reduction—the impact is immediate and easy to measure. In other cases—e.g. better planning and control, increased flexibility—the financial impact is less direct and less certain in magnitude. Nonetheless, the *expectation* of an impact is there.

The discussion so far has considered only the favorable impacts of information systems. It has often been noted (e.g. [1, 2, 9]), that not all impacts of a change to an organization's information system are beneficial. These unfavorable impacts have costs associated with them. Further, they are likely to lead to resistance to the new system, thus reducing or even eliminating the benefits to be derived from it. A complete discussion of these unfavourable impacts is beyond the scope of this paper. Nonetheless, the possibility of this type of impact must be considered in evaluating any new information system, expecially when a choice must be made among alternative systems.

We can summarize this discussion by sketching a normative model of project selection. In order to properly allocate the organization's resources among competing alternative uses, we must be able to identify these alternatives *and* to compare the benefits offered across alternatives. This comparison must include the total package of benefits offered by each project, and should recognize the possibility of unfavorable impacts. Since different projects offer different mixes of benefit types, performing this comparison requires that benefits be quantified. The expectation of a financial impact is the key to this quantification.

In the next section, certain aspects of the MIS project selection process in a major multinational corporation will be examined. The process followed by this corporation will be compared to the normative process outlined above. Where discrepancies are noted, an attempt to explain them will be made. In the final section of the paper, some ways to bring practice more closely into line with the normative model will be suggested.

PROJECT SELECTION AT ALPHA PRODUCTS

Alpha Products is a major corporation, with headquarters in the United States and operations throughout the world. Alpha has a reputation for being progressive in its use of management science techniques (including information systems) and has a long record of experience in this area. Alpha uses a formalized systems life cycle approach similar to those now used by many large organizations. Thus, while not a statistical sample, Alpha should provide data representative of the current state of practice in MIS project selection methodology.

[2] A similar list of difficulties, raised in the context of *computer* selection, can be found in [11].

Data were collected by analyzing written proposals for 71 MIS projects at Alpha. The bulk of these (61) came from three Corporate Office units—Finance, Personnel, and Operations—the remainder coming from overseas units which report to the Corporate Office. The projects ranged widely in size—direct development costs varied from $1000 to nearly $500,000—and in type—including new system development, system extension, program modification, support of production runs, consulting services, special one time analyses, etc. For purposes of analysis, projects were placed into three groups:

(1) Development—the development of a new system or the extension of the *functional* capabilities of an existing system;

(2) Modification—changing report formats, file structures or the specifics of processing in an existing system; and

(3) Support—provision of general MIS consulting services, one-shot reports, etc.

Alpha's written project proposals do not reveal the entire project selection process. They do enable us to examine, however, the estimation of project benefits—the extent to which benefits are quantified, whether quantification is based on a common scale, and the breadth of benefits considered.

As suggested earlier, the benefits claimed for MIS projects are increases in the organization's efficiency—reducing the cost of accomplishing clerical tasks—or effectiveness—enabling better performance of some substantive task.[3] Most of the benefits cited in Alpha's project proposals could be classified as one of these types. However, in a number of instances, another type of 'benefit' was cited—addressing a mandated reporting or recordkeeping requirement. With the inclusion of this third type, all benefits cited could be classified as one of the three types—Efficiency, Effectiveness or Mandated.[4]

The pattern of benefits quantification

Of the 71 project proposals in our sample, 25 (35.2%) make some attempt to quantify (on some scale) the expected benefits (or a portion thereof); the remaining 46 projects either make no attempt to quantify the suggested benefits (38 cases) or list no benefit at all. All proposals in this sample were for projects which had been approved. Thus, a sizable percentage of the projects selected by Alpha for implementation were chosen without documentation of any quantitative estimates of their likely benefits.

To some extent, the lack of quantification is due to the variety of project types and sizes included in the sample (see Table 2). It is apparent that benefits were less likely to be quantified for maintenance projects and small projects than for the development projects and larger projects. Overall, though benefits were quantified for many of the largest projects, for a large number of significant projects this was not the case. Further, we have considered only whether *any* benefits were quantified, not whether the full range of benefits had been examined and their values estimated.

The data in Table 3 address this latter question. The majority of projects claiming any benefit at all list only a single specific benefit or impact on the organization. While this does not prove a failure to consider the full range of system benefits, it is suggestive of one, as most projects have multiple impacts on an organization (consider Knutsen & Nolan's six categories). These data also shed more light on the extent to which benefits are quantified. The tendency to quantify benefits increases—both in the percentage of projects for which at least some benefits are quantified and in the average percentage of the indicated benefits quantified per project—with the number of specific benefits identified. However, even in the categories with the highest degrees of quantification, only about half of the indicated benefits are quantified.

In summary, the pattern of MIS project benefits quantification at Alpha shows that for the majority of projects, no quantification of benefits has been made. This is most apparent on small projects and on projects to support or modify existing systems. The evidence also sug-

[3] The definition of efficiency being used here is a relatively narrow one. Improvements in the Output–Input ratio in *non-clerical* areas will be considered increases in effectiveness for the remainder of the paper.

[4] Some projects cited no benefits at all. While these data could be used for some parts of the analysis, no attempt was made to determine which type(s) of benefit the project provided.

Vol. 7
1979

TABLE 2. BENEFIT QUANTIFICATION[a] BY PROJECT TYPE AND SIZE

	Size[b]			
	Less than $10,000	$10,000–50,000	Greater than $50,000	
Type				
Production support	8*	4	0	12
	(0%)†	(0%)	(—)	(0%)†
System modification	23	4	0	27
	(34.8%)	(25.0%)	(—)	(33.3%)
System development	12	12	8	32
	(8.3%)	(66.7%)	(87.5%)	(50.0%)
	43	20	8	71
	(20.9%)†	(45.0%)	(87.5%)	

* Number of projects.
† Percentage of projects having quantified benefits.
[a] One entry per project, classified as quantified if any benefits for the project are quantified.
[b] Development cost.

gests that consideration of the full range of benefits is not occurring in a large number of Alpha's project proposals. Further, there is no evidence that possible unfavorable system impacts were considered in any of the 71 projects examined. These findings indicate substantial discrepancies between Alpha's practices and the normative model. We can infer that MIS project selection decisions at Alpha are being based on inadequate data, data which cannot assure the best allocation of resources among competing project proposals. And, it is unlikely that this situation is unique to Alpha.

Possible explanations for the pattern observed

Before attempting to prescribe ways to bring practice more in line with the normative model, we should consider some possible explanations for the existing situation. Without some understanding of why things are as they are, it is difficult to suggest meaningful changes.

We will consider four possible explanations for the existing practices:

(1) the economics of analysis,

(2) resource allocation procedures at Alpha,

(3) the political nature of the organizational environment, and

(4) the linkage between observable impacts and financial outcomes.

Economics of analysis. Performing a cost–benefit analysis is time consuming and costly. For some projects, especially small ones, the possible benefits to be derived from a rigorous examination may be smaller than the cost of examination. For such projects, it would not make economic sense to conduct an analysis. Thus, the substantial number of small projects included in the sample may provide a partial explanation for the existing pattern.

TABLE 3. NUMBER OF BENEFITS LISTED BY PROJECTS

Number of specific benefits claimed	Number of projects	Percent of projects	Percent of projects with *some* benefits quantified	Percent of specific benefits quantified
0	8	11.3	0	0
1	32	45.1	28.1	28.1
2	15	21.1	46.7	36.7
3	9	12.7	44.4	22.2
4	3	4.2	66.7	50.0
5 or more	4	5.6	100	51.9
	71	100		

TABLE 4. BENEFIT QUANTIFICATION BY BENEFIT TYPE WITHIN FUNCTIONAL AREA

	Corporate finance	Corporate personnel	Corporate operations	
Mandated	2*	13	0	15
	0%†	0%	—	0.0%)†
Efficiency	5	15	7	27
	20%	46.7%	71.4%	(48.1%)
Effectiveness	7	11	8	26
	0%	0%	62.5%	(19.2%)
	14	39	15	68[a]
	(7.1%)†	(17.9%)	(66.7%)	

* Number of benefit classes listed by projects.
† Percentage of benefit classes listed having quantified benefits; a class was considered quantified if *any* benefit of that class was quantified.
[a] Total number of benefit classes listed in the 61 projects from corporate office units.

Resource allocation procedures. Alpha allocates a major portion of available system development resources to organizational units when annual budgets are drawn up. Each unit has the responsibility to manage the resources allocated to it, and the allocation of resources to specific projects is accomplished in the local environment. The manager of an organizational unit may feel that he (or she) has enough understanding of that environment to make allocation decisions without rigorous analyses of project impacts. Again, this may explain in part the observed pattern.

The political nature of the environment. Organizations are acknowledged to be political entities [13]. As a result, actions taken are not always based solely on economic criteria. Other criteria may supplement, or even supplant, economics in the project selection decision. If non-economic criteria play a dominant role in project selection, there is little reason to conduct formal economic analyses. No formal data were collected at Alpha to test the extent to which political (non-economic) factors influence project selection decisions. However, the large number of projects selected at Alpha without formal economic justifications is consistent with political behavior.

Linkage to financial outcomes. Some of the impacts of MIS projects can be linked quite readily to financial outcomes; for others the linkage is not so straightforward. The data from Alpha provide two examples of this phenomenon—the contrast between efficiency and effectiveness benefits, and differences in the degree of linkage as between different functional areas. In the former case, better models generally exist of the relationship between efficiency improvements (i.e. changes in data handling operations) and financial outcomes than of the relationship between effectiveness improvements and financial outcomes. The data (Table 4) bear this out. There is a markedly higher incidence of quantification of efficiency benefits than of either of the other types.

Across functional areas, generally the models linking activity to financial outcomes are better for functions close to the organization's outputs (e.g. production or distribution) than for functions which are distant from those outputs (e.g. finance or personnel). The more explicit the model for a functional area is, the more likely it is that the value of changes in the information system for that area can be estimated. Table 4 indicates that Finance and Personnel show markedly lower tendencies to quantify project benefits than does Operations, and this is consistent with the linkage hypothesis.

This difference across departments also appears when we look at the three classes of benefits separately. Operations is less likely than the staff departments to claim a system as mandated. And, in both the efficiency and effectiveness classes, Operations shows substantially higher percentages of benefits quantified than do the Personnel and Finance departments. While the linkage hypothesis might explain the difference for effectiveness benefits, improvements in data processing efficiency should be no more difficult to quantify in Personnel than in Operations. Perhaps some sort of 'halo effect' is in operation; the practice of not quantifying benefits in areas where models are fuzzy, spills over to the efficiency area where good models do exist.

Vol. 7
1979

These four practices may well account for much of the incomplete analysis of project benefits observed at Alpha. Before suggesting ways to change the project selection process, we must consider the purposes served by the four practices.

Not performing a rigorous analysis where the cost of analysis is high and that of the project low is sensible. Submitting all projects to thorough analyses could increase costs far more than it would likely increase benefits. There is a danger, however, that this procedure will result in the bulk of available resources being allocated to low cost, and likely low benefit, projects, leaving inadequate resources for larger projects with greater potential.

By placing decision and control responsibility in the hands of the local unit manager, lump-sum allocations garner all of the benefits normally suggested for organizational decentralization [10, Chapter 24]. However, by giving individual organizational units uncontested claim to a certain quantity of resources, this procedure introduces a disincentive to rigorous analysis of project benefits. Such analyses could show that one unit's projects were not as beneficial as those in other units, and this could lead to a loss of resources. If the analyses are never made, inter-unit comparisons cannot be made, and each unit is assured of keeping its pre-allocated share of the pie.

Political behavior in organizations is a fact of life. It would be difficult, and probably undesirable, to eliminate political criteria from the project selection decision. While the use of non-economic criteria is not by itself bad, the impact of that use may be to drive out hard analysis. Without analysis, it is difficult to assess how well the allocation of resources to projects matches the organization's objectives. Economic analyses should be one input to the admittedly political decision making process.

In the case of the linkage hypothesis, the question is not so much whether it is desirable, but rather whether it is inevitable. Are staff functions so remote from the business of the organization that the impacts of changes in their methods are essentially non-quantifiable? The ten project proposals from Alpha's overseas divisions suggest this is not so. All of the projects proposed by these divisions listed quantified benefits. All efficiency benefits are quantified (7 projects), as are half of the effectiveness benefits (6 projects). Of these ten projects, nine are in the finance area and one is in operations. Two of the financial systems are justified solely on the basis of quantified effectiveness type benefits. Thus, it is *possible* to estimate the impact of effectiveness improving information systems in areas remote from the organization's product.

In summary, of the four practices suggested as possible explanations for the pattern of incomplete project analysis observed at Alpha:

(1) two—economics of analysis and lump-sum allocations—can serve useful purposes but can also be dysfunctional;

(2) one—the linkage hypothesis—is probably quite common, but serves no useful purpose and can be avoided; and

(3) one—the political environment—is a fact of life sometimes used as an excuse for failing to perform an analysis.

The next section of this paper considers an approach to modifying the project selection process to bring it closer to the normative model. These modifications directly address the problem of linking information system changes to financial outcomes, and may impact the economics of analysis by making good analysis less costly. They do not address the problems posed by lump-sum resource allocations nor by a political environment which chooses to avoid explicit analysis.

SOME DIRECTIONS FOR CHANGE

One key requisite for improving benefits estimation is the development of a meaningful taxonomy of benefit types. The purpose of this taxonomy is to help identify the full range of project benefits and to enable their measurement. To achieve this end, the taxonomy should

(1) be comprehensive, covering all types of system benefits,

(2) be concise, lest it be too cumbersome to be of use, and

(3) provide guidance for the benefit measurement and comparison processes.

TABLE 5. CLASSIFICATION OF CITED INFORMATION SYSTEM BENEFITS

	Benefit class	No. of cases	% of cases	% quantified
Mandates	1. Record or report required information ('Mandated')	14	11.3	0.0
Organizational processes	2. Reduce information processing or handling costs ('Efficiency')	47	37.9	70.2
	3. Improve asset utilization and resource control	12	9.7	83.3
	4. Improve planning process	1	0.8	100.0
	5. Increase organizational flexibility	4	3.2	0.0
	6. Promote organizational learning	0	—	—
Information characteristics	7. Provide greater accuracy in clerical operations, fewer errors	20	16.1	0.0
	8. Provide information on a more timely basis	6	4.8	0.0
	9. Provide new/more/better information	16	12.9	6.3
	10. Other	4	3.2	50.0
	Total	124	100	

Vol. 7
1979

The taxonomies introduced so far in this paper meet the first two criteria, but fail on the third.

The following taxonomy is proposed as a starting point:

(1) Record or report required information (Mandated);

(2) Reduce information processing or handling costs (Efficiency);

(3) Improve asset utilization and resource control;

(4) Improve planning process;

(5) Increase organizational flexibility;

(6) Promote organizational learning;

(7) Provide greater accuracy in clerical operations, fewer errors;

(8) Provide information on a more timely basis; and

(9) Provide new/more/better information.

This taxonomy appears to meet the three criteria suggested above. Clearly, it is concise; it includes only nine categories of benefits. We can test the comprehensiveness of these categories using data from Alpha. The benefits listed in the 71 project proposals were classified according to the proposed taxonomy. Table 5 shows the results of this classification, and samples of the benefits falling into each category are presented in the appendix. All but four of the 124 specific benefits listed in these proposals could be placed in one of the nine categories. The remaining four could not be categorized because they differed in level of analysis from the categories in the taxonomy (to be discussed below).

While the proposed taxonomy was adequate for classifying almost all statements of benefit included in Alpha's MIS project proposals, it is doubtful that the taxonomy is exhaustive as it stands. First, only beneficial impacts have been included. Categories representing unfavorable impacts might be added; but the analysis of these impacts and their classification is beyond the scope of this paper. Second, additional categories of beneficial impacts might prove necessary, and these could be added as the need for them is shown.

Finally, we can consider the guidance this taxonomy provides to the measurement and comparison processes. Information system impacts can be analyzed at a number of levels.

The highest level is the system's impact on organizational outcome variables—sales revenue, customer satisfaction, profit contribution, etc. These are the *ultimate impacts* of the system. They do not follow from it directly and immediately, but rather result from changes to *organizational processes*—e.g. planning, control, learning. These processes represent a second level of analysis, and are included in the taxonomy as benefit classes 2–6. There is a third, still lower, level at which information system impacts can be analyzed: the immediate changes to the *information produced* by the system and available to the organization. These changes, represented by benefit classes 7–9, affect the organization only through the processes listed above them.

Eight of the nine benefit classes included in the taxonomy fall into these latter two groups—changes to organizational processes and changes to information produced. Ultimate impacts were not included; first, because the number of unique ultimate impacts is too large, and second, because they can be determined once the process changes are identified. The ninth benefit class included in the taxonomy—mandated changes to information systems—forms a group by itself.

The final column in Table 5 shows the degree to which benefits were quantified within each of the classes. The bulk of the quantification occurred in those classes representing changes to organizational processes. This is not a coincidence, and it provides useful insight. Changes to processes are the link between changes to information and organizational outcomes. It is only once we understand *how* the new information will be used that its value can be estimated. Thus, efforts to quantify benefits should focus on the changes in organizational processes which will result from changes to information systems.

It should be possible to develop estimates for all benefits in the process change category. The degree of certainty will vary from class to class, and the ordering of these classes is meant to approximate decreasing certainty of estimates. At one extreme, the impact of increased data processing efficiency can be estimated with relative certainty. Toward the other extreme, the impact of increased organizational flexibility is far less certain. But, this does not mean that it cannot be estimated. There is evidence to suggest that managers have fairly good implicit models of the processes that affect their organizations [4]. We should be able to tap into these models to develop at least a range (e.g. best case–worst case scenarios) of estimates for the value of these less certain process changes.

Changes to information characteristics are, by themselves, of no value, and it does not make sense to attempt to estimate values for informational changes as such. Rather, consideration should be given to how these changes will affect organizational processes. The impact of those process changes can then be estimated.

In the case of mandated changes, there really is no reason to estimate the value of the change. Where a particular change truly is mandatory, emphasis should be placed on accomplishing the change at the lowest cost. Care must be taken, however, to assure that classifying a change as mandatory is not simply a device used to avoid rigorous analysis of a project.

In summary, the taxonomy helps to guide the measurement process by focusing it on those types of benefits which can be estimated, process changes. Other benefit types should first be converted to a process change, and then an estimate can be made.

One problem remains. Since uncertainty about the value of a benefit will vary from class to class, meaningful project to project comparisons cannot be made simply by summing all benefits for a given project and comparing totals across projects. That would imply that the sole concern was the expected value of the benefits without regard to risk. Risk and return must both be considered, and the organization's risk preference must enter into this consideration.

At least two general approaches for comparing projects having differing mixes of benefits (and hence, uncertainty) seem possible. One is to apply 'discount factors' to expected benefits according to the degree of uncertainty they present, and then to sum across benefits. The more uncertain the benefit, the larger the discount would be. Scoring models (such as in [8]) are representative of this general approach. The problems with this technique are (1) it likely understates the relative value of a project with a preponderance of benefits in the more uncertain categories, and (2) it may

be difficult to determine discount factors which reflect the organization's risk preferences over a range of situations (e.g. type of benefit, size of the minimum or maximum benefit).

A second approach is to present each project as a vector, showing for each benefit category minimum, maximum, and most likely values. This addresses the two problems raised above, but presents a new one. The individual decision maker must now make the risk–return trade-off, and there is evidence which suggests that individuals who see only a subset of the organization's activities are more risk averse than are top managers of the organization [5]. No solution is offered at this point, but both approaches seem to be worth trying.

SUMMARY AND DIRECTIONS FOR FURTHER RESEARCH

This paper has suggested that project selection is an important, often overlooked step in the MIS development process. Effective project selection decisions require consideration of all benefits to be derived from a project and measurement of these benefits on comparable scales. Current practice often involves estimating those classes of benefits which are easiest to measure and about which there is little uncertainty, and describing other very real benefits as intagnible. In fact, these latter benefits *are* tangible; however, they cannot be measured as easily or with as great certainty as can the former group.

A taxonomy to help identify the full range of benefits was presented, and the outline of an approach to benefits measurement was sketched. The next step must be a test of this taxonomy and refinement of the measurement approach. Among the questions requiring answers are:

(1) will application of the taxonomy lead to more complete identification of project benefits?

(2) is the taxonomy rich enough to capture the full range of benefits?

(3) can changes to information characteristics be linked to organizational process changes?

(4) how can estimates differing in uncertainty best be combined and compared?

(5) can this approach work in an inherently political environment? and

(6) what impact, if any, does following this approach have on project selection decisions—that is, will they improve?

[5] Number of specific statements in this class in the sample of 71 projects at Alpha Products.

ACKNOWLEDGEMENT

The author wishes to thank Martin Starr, Jon Turner and Miklos Vasarhelyi for their helpful comments on earlier versions of this paper.

REFERENCES

1. Argyris C (1971) Management information systems: the challenge to rationality and emotionality. *Mgmt Sci.* **17**(6), B275–B292.
2. Argyris C (1977) Organizational learning and management information systems. *Acctg, Orgns & Soc.* **2**(2), 113–123.
3. Carlson ED (1974) Evaluating the impact of information systems. *Mgmt Informatics* **3**(2), 57–67.
4. Gorry GA (1971) The development of managerial models. *Sloan Mgmt Rev.* **12**(2), 1–16.
5. Hammond JS III (1967) Better decisions with preference theory. *Harvard Busin. Rev.* **45**(6), 123–141.
6. King JL & Schrems EL (1978) Cost–benefit analysis in information system development and operation. *Comput. Surv.* **10**(1), 19–34.
7. Knutsen KE & Nolan RL (1974) On cost–benefit of computer-based systems. In *Managing the Data Resource Function* (Ed Nolan RL), West Publishing Co, St. Paul, USA.
8. Lucas HC Jr & Moore JR (1976) A multiple-criterion scoring approach to information system project selection. *INFOR* **14**(1), 1–12.
9. Mumford E & Banks O (1967) *The Computer and the Clerk*. Routledge & Kegan Paul, London, UK.
10. Shillinglaw G (1977) *Managerial Cost Accounting*, 4th ed. Richard D Irwin, Homewood, Illinois, USA.
11. Timmreck EM (1973) Computer selection methodology. *Comput. Surv.* **5**(4), 199–222.
12. Tucker CC (1978) From the chairman. *Data Base* **9**(3), 2–3.
13. Tushman ML (1977) A political approach to organizations: a review and rationale. *Acad. Mgmt Rev.* **2**, 206–216.

ADDRESS FOR CORRESPONDENCE: *Michael J Ginzberg, Esq, Associate Professor of Business, Graduate School of Business, Columbia University, New York, NY* 10027, *USA.*

APPENDIX
Samples of Benefits Specified in Alpha Products' MIS Project proposals

I. Record/Report Required Information ('Mandated') —$n = 14$[5]
   1. comply with court order
   2. top management committee mandate

3. to comply with ERISA
4. audit department recommendations
5. company policy
6. tax department request

II. Reduce Information Processing/Handling Costs ('Efficiency')—$n = 47$
1. relieve payroll department manual effort
2. reduce reruns
3. reduce time spent researching and correcting errors
4. man-hour time savings over prior method
5. reduce number of Master Files created
6. savings in cost of report mailing
7. eliminate double data entry and manual updating
8. manpower savings through automation of analysis
9. eliminate maintenance on existing system
10. reduce training needs
11. improved efficiency in assembling and distributing bills of material
12. simplify user procedures for processing requests
13. save time and effort on future system modification

III. Improve Asset Utilization & Resource Control—$n = 12$
1. reduced transport delays
2. reduction in equipment maintenance costs and costs due to equipment failure
3. better allocation leading to increased product value
4. better control of inventory
5. better use of reworked goods through better information about product characteristics
6. permit more effective control of projects by managers and analysts

IV. Improve Planning Process—$n = 1$
1. better composition of product lines

V. Increase Organizational Flexibility—$n = 4$
1. flexibility to handle special mailings
2. system easily modifiable as operating environment changes
3. system will allow quick and accurate response to crises
4. flexibility to adapt system to changing management needs for information

VI. Promote Organizational Learning—$n = 0$
None

VII. Provide Greater Accuracy in Clerical Operations; Fewer Errors—$n = 20$
1. assure proper calculation of premiums
2. better control over price changes—fewer failures to update billing price to current level
3. better recovery of customs duty paid
4. improve audit, control, and coordination of benefit plan payments
5. provide management with current reports and insure data integrity
6. verify aggreement with insurance company records
7. elimination of inter-system inconsistencies
8. identify errors before check printing

VIII. Provide Information on a More Timely Basis—$n = 6$
1. enable running of tests on regular basis without need for analyst involvement
2. provide accurate information more quickly so that decisions can be made sooner (production and inventory decisions)
3. complex analyses can be made available in more detail, more quickly, and more accurately

IX. Provide New/More/Better Information—$n = 14$
1. file needed to produce retirement projections
2. system easier to use and more understandable
3. more effective provision of information required for purchasing and vendor delivery scheduling
4. able to do more analysis of operations data and determine financial effects
5. improve communication with beneficiaries
6. analyze impacts of regulations on existing systems
7. make needed format changes

X. Other—$n = 4$
1. prevent possible recurrence of a past problem
2. better technical assistance to customers leading to higher profit contribution
3. increased profit contribution
4. provide needed support

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 539 to 543
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0539$02.00/0

Evaluation of Management Information Systems

JPC KLEIJNEN

Katholieke Hogeschool, Tilburg

(*Received January*, 1979)

The economic evaluation of Management Information Systems may be based on the following theories and techniques: Control Theory, System Dynamics, (discrete-event) simulation, and gaming. Applications of these approaches are summarized. Advantages and disadvantages of the various approaches are presented.

1. INTRODUCTION

TRADITIONALLY the emphasis in computer science and information science (called 'informatica' in Europe) has been on the evaluation of the *technical* performance of computer systems, measured by such criteria as throughput, response time, CPU (Central Processing Unit) utilization, (see, for instance, [7]). Only a few attempts have been made to model the *economic* performance of computerized information systems—or MIS (Management Information Systems)—measured by criteria like profits, sales, inventories etc. In this paper we shall evaluate economic evaluation studies based on System Dynamics models and 'discrete' simulation.

2. CONTROL THEORY

Control Theory highlights the role of feedback and feedforward information. Negative feedback means that information on a variable is compared to some desired goal value, and in the case of deviation between the actual and the goal value, corrective action is taken. Feed forward means that predictions or forecasts are made, such as sales forecasts. The advantage of Control Theory over techniques such as simulation is that it provides more general conclusions. The disadvantage of Control Theory is that it must introduce drastic simplifications in order to keep the mathematical problems within manageable limits. Hence its models are usually continuous, i.e. a very aggregated view of the world is maintained. The models are also linear in order to simplify analytical solutions (using Laplace transformations). Steady-state solutions are relatively easy to obtain so that transient behavior tends to be neglected. Deterministic models are simpler to solve than stochastic models. Hence bias but no stochastic noise was examined in the study by Politzer & Wilmès [9] which we shall discuss below.

Delays are modeled through, e.g. *first order exponential delays*. The effects of such delays are illustrated in Fig. 1: In the upper part the input shows a temporary increase (pulse) and in the lower part the input shows a permanent increase (ramp function). The reaction of the output depends on the order of the delay function, as the figure illustrates. These delay concepts correspond with a *continuous* view of the world as is maintained in electrical engineering. For a critical discussion of such delay functions (also used in System Dynamics) we refer to Riethmüller & Schreiber [10]. In Kleijnen [6] several sources of delay in MIS are discussed, such as update interval (say, weekly versus

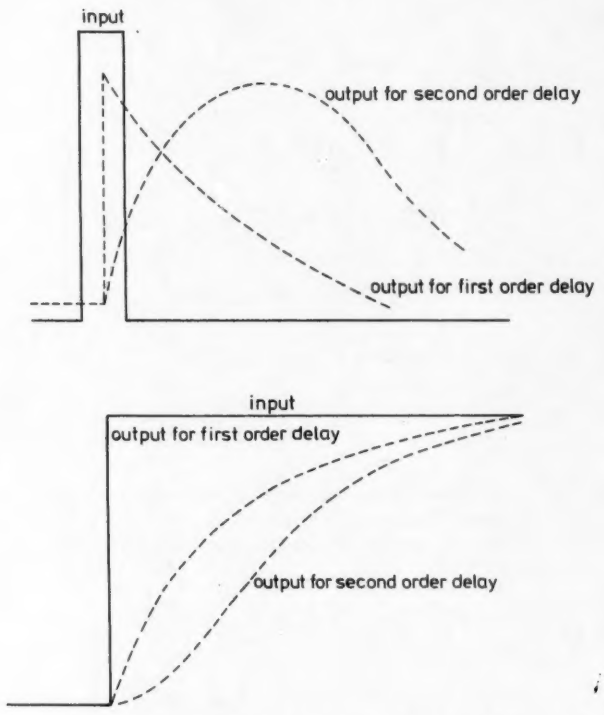


FIG. 1. *Delays of first and second order*

monthly processing), retrieval delay in real-time-systems, decision delay (time to reach a decision and to affect the physical system) etc. These sources of delay would require non-continuous, non-linear, stochastic models if we want a realistic model. So Control Theory is far removed from the mental models used by System Analysts and most users in data processing. It remains possible that at higher levels of decision-making a more aggregated view as proposed by Control Theory, becomes more appropriate.

Politzer & Wilmès [9] applied Control Theory to the evaluation of a MIS. They restricted their study to a classical model for planning inventories and production originated by Holt *et al.* [5] and known as the HMMS model. This model assumes simple quadratic cost relationships. The authors derived a steady-state solution based on a continuous version of the HMMS model, and they studied the effects of (exponential) 'delay' and bias. Gross value was found to react asymmetrically to multiplicative bias, i.e. overestimation had a different effect compared to underestimation. Delay was found to have an optimal value larger than zero, i.e. on-line systems do not necessarily yield higher gross benefits. However, remember that these conclusions were derived under a number of restrictive assumptions!

If the Control Theoretic model cannot be solved analytically then we may resort to simulation, especially System Dynamics simulation.

3. SYSTEM DYNAMICS

System Dynamics (SD) or Industrial Dynamics has as its philosophy that each socio-technical system can be looked upon as a dynamic system with (negative) *feedback*. Since it takes time to collect and process data on the output, to compare the resulting information to the norm, and to adapt the decision variables, *time lags* are created which make the model dynamic. Delays in information transmission may cause fluctuations within the company, as opposed to externally caused business cycles. The computer language that has been developed especially for this SD approach is DYNAMO, but other languages can also be

Vol. 7
1979

used, e.g. FORTRAN or the simulation language CSMP.

There has been considerable controversy about the usefulness of the SD methodology. SD may tackle any problem, ranging from some subsystem within a company to such world problems as world energy resources. The approach does not pretend to give exact numerical predictions but it does show the dynamic properties of a system [1]. Thissen [13] examined a variety of techniques for improving the understanding of complex SD models. Nordhaus [8] emphasized that in SD relationships and variables are not based on empirical data—in contrast with traditional econometrics—while changing an assumption may drastically affect the model's outcome. Compared to other simulation approaches, SD has the advantage of being more than a technique. However, the very advantage of being a framework means that the modeler has to adopt a world view (with levels and rates) that is not spontaneous, but requires learning (or indoctrination?). For instance, a discrete–event simulation of a queuing system is a direct representation of the spontaneous, intuitive way such a system would be described by 'anybody'. Unfortunately, such a discrete–event description is too detailed for 'large' systems. Note that if one is willing to adopt the SD view of the world, mastering the SD technique needs practice in the *art* of model building.

Of special interest to us is the concept of *delay* in SD. As in Control Theory SD uses delays of first order, second order, etc; see Fig. 1 above. We find a discrete view more compatible with the MIS world view (see section 2).

Though a great many studies have been performed using the SD methodology, only a few of these studies concern the explicit study of information systems.

Swanson [12] investigated the effects of error, distortion (bias), delay, and sampling (update interval). He modeled the control of manpower, production, and inventory, based on information on sales and inventory. Swanson found that delay had the largest effect, while error had the smallest effect. His conclusions, however, are not based on solid statistical methods: only two simulation runs per case with no interactions measured.

At Philips Industries a series of studies was performed under the acronym IPSO: Initiating Production by Sales Orders [3, 14]. This studied the effects of delays in information and examined the scope of the IS. In a chain of production processes—or more generally, a network of production processes—a process may be provided with information on the demand from the next process. Alternatively, this information may be augmented with marketing information, i.e. the demand by the ultimate customers of the company. Extending the scope of the data base so that besides 'local' data 'global' data are also available, was found to dampen the oscillations in the company's inventories and activities. Note that this series of studies combined a number of techniques and ideas ('models'). The models were strongly influenced by System Dynamics. The simulation experiments were combined with analytical results derived for simplified models, e.g. Control Theory models. Analysis suggested which SD models to simulate, and simulation results were used to test analytical approximations.

In conclusion, SD is an extremely popular method. It is much more than a technique; it is a world view. SD is best known for its study of large scale systems: World Dynamics, Urban Dynamics, etc. For such large systems an aggregated view seems necessary. At such high levels of aggregation ('the' pollution, 'the' world consumption) continuous models may be adequate: flows, rates, exponential delays. However, at the level of MIS we prefer more detailed models with discrete events, aperiodic decisions, and realistic representations of information attributes. A technique permitting such a level of detail is simulation together with its extension, gaming. Note that the derivation of adequate decision rules (policies) is assisted by SD modeling. Especially when the MIS is integrated (e.g. coupling the production and the inventory system), then large systems must indeed be managed.

4. SIMULATION

The *structure* of simulation models can be different in the following respects. The system may be modeled by *difference equations* so that at *equispaced* points of time the state of the system is computed. An example is a corporate simulation model based on balance and profit & loss account relationships. Slightly more

complicated difference equations are used in System Dynamics: a set of recursive difference equations with variable coefficients (rates). In *discrete event* simulation the model is a most detailed representation of the real-world system in so far as individual events are represented. Many languages have been developed for this approach, e.g. GPSS and SIMSCRIPT. In this section the term 'simulation' will exclude SD simulation models.

Simulation permits the most realistic models. Moreover, the user—the client of the model builder—can better understand a simulation model. A disadvantage is that it may take much effort to build and run a simulation model, while the responses are, strictly, valid only for the specified values of the input variables and parameters. In Kleijnen [6], however, we have explained how statistical techniques can be used to generalize simulation results. Next we shall present an example of a simulation study on the value of information.

Boyd & Krasnow [2] performed an interesting *discrete-event* simulation study on the timeliness of information in the context of a hypothetical manufacturing company. They investigated periodic decision-making, tactical decisions being made less frequently than operational decisions. They also varied the processing delay (turnaround time). The delays affected the speed with which decisions could react to changes in the environment such as changes in demand. This lead to fluctuations in both profit and physical variables like inventory. Reductions in the various delays improved profit significantly.

Note that Welke [15] criticized the *ad hoc* character of simulation studies such as made by Boyd & Kranow [2]. He proposed a general framework for the simulation modeling of information systems. His framework supports a user's view, as opposed to the traditional technical data processing view. The framework can be applied to a particular system in order to generate a descriptive model, which can next be evaluated using a discrete–event simulation package. We refer to Welke [15] for more details.

5. GAMING

In gaming or man–machine simulation one or more input variables is determined by human participants or 'players'. Let us briefly compare gaming to pure computer simulation including System Dynamics simulation. The relative advantage of gaming is its *behavioral realism*: Decision-makers may have multiple—possibly vague—goals. They may use heuristic—possibly inconsistent—decision rules. They may form competing or colluding teams. They may show learning behavior or fatigue and boredom, and so on. The disadvantages of gaming relative to pure machine simulation, are lack of *experimental control* and certain practical nuisances. If the purpose of the game is to reach general conclusions rather than to train people, then human participation creates much noise. For instance, a player can show inferior performance not because he is supplied with poor information but because he lacks 'feeling', gets bored, is confronted with superior opponents, etc. General conclusions require repeated runs with the game. But repeated runs are expensive or may even be impossible because of lack of players and computer time. Moreover, repeated runs are not completely comparable since the players show learning, fatigue etc.

This brief discussion was not meant to decry gaming as a useful research instrument, but to warn against possible pitfalls. A general reference on gaming is Shubik [11].

The '*Minnesota*' *gaming experiments* were performed at the University of Minnesota's MIS Research Center during the period 1970–1975. In these experiments relationships were investigated among

(i) the information system itself;
(ii) the decision-maker: psychological characteristics, experience;
(iii) the decision environment: operational versus strategic, etc.

Dickson *et al.* [4] summarized nine different gaming experiments. The experiments taken together, showed that effects did occur when changing the experimental factors, i.e. information system, decision-maker, environment. Unfortunately, no simple general relationships could be formulated.

6. CONCLUSION

The evaluation of the economic benefits of MIS forms an underdeveloped area. In order

Vol. 7
1979

to make progress in this area, we need *theory plus tools*. Theories provide framework, clues, guidelines, ideas, etc. Tools provide numerical results. In the context of MIS relevant theories seem to be: Control Theory, System Dynamics, Bayesian Information Economics. Useful tools are simulation, gaming, and—more generally—mathematical modeling and statistics. A new management style, fostered by more scientifically trained managers, might stimulate the acceptance of formal models. This paper surveyed a number of formal theories, models, and techniques. Their practical application remains as the major task.

REFERENCES

1. ANSOFF HI & SLEVIN DP (1968) An appreciation of industrial dynamics. *Mgmt Sci.* **14**(7), 383–397.
2. BOYD DF & KRASNOW HS (1963) Economic evaluation of management information systems. *IBM Systems Jl* **2**, 2–23.
3. BRAAT JJM (1973) The IPSO control system. *Int. Jl Prod. Res.* **11**(4), 417–436.
4. DICKSON GW, SENN JA & CHERVANY NL (1977) Research in management information systems: the Minnesota experiments. *Mgmt Sci.* **23**(9), 913–923.
5. HOLT CC, MODIGLIANI F, MUTH JF & SIMON HA (1960) *Planning Production, Inventories and Work Force*. Prentice-Hall, Englewood-Cliffs, New Jersey, USA.
6. KLEIJNEN JPC (1980) *Computers and Profits: Quantifying Financial Benefits of Information*. Addison-Wesley, Reading, Massachusetts, USA.
7. KOBAYASHI H (1978) *Modeling and Analysis: an Introduction to System Performance Evaluation Methodology*. Addison-Wesley, Reading, Massachusetts, USA.
8. NORDHAUS WD (1973) World dynamics: measurement without data. *Econ Jl* **83**, 1156–1183.
9. POLITZER JP & WILMÈS P (1977) Optimal Design of Management Information Systems through Control Theory. Working Paper **77-5**, Faculté d'Administration, University of Sherbrooke, Québec, Canada.
10. RIETMULLER CE & SCHREIBER BD (1977) Delays in socio–economical systems dynamics models. In *Simulation '77, Proceedings of the International Symposium* (Ed. HAMZA MH), Acta Press, Anaheim, California, USA.
11. SHUBIK M (1975) *The Uses and Methods of Gaming*. Elsevier, New York, USA.
12. SWANSON CV (1971) Evaluating the Quality of Management Information. Working Paper, Sloan School of Management, Cambridge, Massachusetts, USA.
13. THISSEN WAH (1978) Investigations into the Club of Rome's World 3 Model; Lessons for Understanding Complicated Models. Doctoral dissertation, Technical University, Eindhoven, Netherlands.
14. VAN AKEN JE (1978) On the Control of Complex Industrial Organizations. Martinus Nijhoff Social Sciences Division, Leiden, Netherlands.
15. WELKE RJ (1975) Information System Evaluation—a "Non-Bayesian" Approach. Faculty of Business, McMaster University, Hamilton, Ontario, Canada.

ADDRESS FOR CORRESPONDENCE: *Dr JPC Kleijnen, Economische Faculteit, Katholieke Hogeschool, (Tilburg University), Postbus* 90135, 5000 *LE Tilburg, Netherlands.*

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 545 to 551
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0545$02.00/0

Assessment of Fifteen Years of Urban Modeling

WILBUR A STEGER[1]

CONSAD Research Corporation, Pennsylvania, USA

(*Received February* 1979; *in revised form March* 1979)

The purpose of this paper is, through an 'ethnographic' approach, to review where the urban and regional modeling field has been since its inception, regarding analytical inputs to policy analysis, policy development, policy assessment etc. Assessed are: What progress in understanding, explanation, evaluation has there been? Do we know a *little* or a *lot* more after 15 years than we did in 1962–1963? More than anything—given our often overly ambitious aspirations, false starts, and imprecise objectives: Can representative members of the urban/regional modeling profession be more helpful to policy makers today than we were in the mid-sixties and early seventies? The paper reviews the phases through which this particular modeling field has achieved its current state of maturity, relative calm, and usefulness, and concludes with a discussion of how we could do even better.

INTRODUCTION

THERE ARE, literally, hundreds of urban 'land use' and regional development analytic models in various stages of development and use today. Given that these are continuously critiqued in detail, elsewhere, I thought it would be worthwhile to perform a 'broad-brush' review, one that touches upon some major phases of design and use since the early sixties when these models first burst upon the urban analysis scene. My purpose, then, is to explore in this paper:

- Where have we come, in the urban policy analysis/modeling 'life cycle'? What phase are we in, both 'psychosocially' and technically? Where are we going?
- How can we assess where we are today? How 'good' are the analytic models? What can we do to better evaluate where we are?

[1] Paper presented at the Urban and Regional Planning Session, TIMS/ORSA, New York City, USA, 1–3 May, 1978.

TAKING STOCK OF THE RECENT PAST

Part of the research I did for this paper consisted of reviewing past reviews, including my own of more than 10 years ago [20] and at least a dozen earlier reviews of urban and regional models and their application [7, 9, 11, 12, 14, 15, 18, 19, 21]. Then we were focusing, and I believe appropriately at the time, more on technical issues of 'How do we *build* these analytical tools?' than 'How do we *use* them?', or 'How do users *assess* using analytical tools vs other (less analytical) approaches?' Today, however, emphasis *is* on these issues, and quite appropriately so. My focus, here, will be on the *policy analysis* use of models, primarily, leaving other more functionally-oriented models to other reviewers.

So, to begin: Where has the urban and regional modeling field been since its inception, regarding analytical inputs to policy analysis, policy development, policy assessment etc.? What progress in understanding, explanation, evaluation has there been? Do we know a *little* or a *lot* more after 15 years than we did in 1962–1963? More than anything—given our

often overly ambitious aspirations, false starts and imprecise objectives: Can representative members of the urban/regional modeling profession be more helpful to policy makers today than we were in the mid-sixties and early seventies?

During this 15-year period, modelers have exhibited a certain schizophrenic attitude toward policy making and policy makers. (Note should be taken of my 'ethnographic' approach to the subject: I am reporting, entirely undocumented by specifics, on what I have perceived over the years, by and large, and generally *feel* to be the case. To attempt to perform this 'trend analysis' by more detailed analytic classification and quantification would be both interesting and, perhaps, even revealing: I leave that project to others).

What I want to do, in this paper, is to briefly recapitulate where I believe modeling *has been* in working with 'policy' matters. I have chosen to include within that purview all matters concerning 'the art or science of the conduct of public affairs... leading to a definite course or method of action selection from among alternatives and, in the light of given conditions, to guide and usually determine present and future decisions' (Webster's *Third New International Dictionary*)[2]. Without meaning intentionally, to slight the advances of what has come to be known as 'policy sciences' in recent years [5, 13] and risking offending those who do not use the word they are defining *in* their definition, I like Ken Kraemer's introductory discussion about 'policy analysis' [10]:

> A city council awakens to the fact that various past municipal policies have tended to discriminate against the poor and against blacks and other minorities in the community but is at a loss as to how to reverse the situation.... Each of these policy makers and groups senses that things are not as they should be. They would like to make improvements, but they feel that they do not really know what the problem is—that what they now perceive as a problem may be only a symptom or a part of a problem far greater in scope and complexity. They do not know how to find out what they do not know.... These policy makers have some ideas on possible solutions to the problem as they now perceive it, but they would like other ideas and insights.
>
> They usually find it difficult to think of alternative solutions realistically different from those presently in use, yet they feel that better alternatives could be found. But finding them is in itself a problem.
>
> Although the policy makers may have some ideas about what might be done, they are usually uncertain as to the consequences of taking action, especially in view of their limited understanding of the problem. At the same time, they feel that if they fail to act immediately the problem may get worse.

So 'policy-making' is what policy-makers do—and, in the public sector, it takes as many forms as there are policy makers.

Phases of growth

Modeling's schizophrenic attitude toward this pragmatically put state-of-affairs has been manifested in the various phases of: groping and exuberance abetted by naïveté, in its earliest years; enthusiasm combined with a natural and romantic optimism of the mid-later 60's; an inward-looking and questioning retrospection in the early 70's; and some frontal attacks, along a number of fronts, in more recent years, to improve the effectiveness of structured modeling-produced information in policy-making.

The early years. Nothing seemed to escape the attention of the early modeling pioneers. Whether the functional area was economic development, physical planning, housing choices, the environment, fiscal and budgetary authority, no matter: the world of models would probably have a role, an important role. In the years 1963–1967, modelers ranged over a wide range of urban and regional policy issues and their accompanying information systems. Few subnational public sector issues escaped attention. The purview included areally diverse, wide-ranging multistate regional development policy matters (e.g. should we pursue a 'worst-first' approach to aiding underdeveloped regions), through the newly developing metropolitan area regional planning agencies, to cities with the myriad problems not yet even partially resolved 15 years later, to neighborhood city subareas and the first glimmer that these could be dealt with *analytically* through analytical models.[3]

I would characterize these first years as 'exuberant' rather than as 'cocky' or 'self-indulgent'. Modelers were groping for their roles, responsibilities and, if you will, the very 'discipline' which could unite them in the loosely-knit federation of urban analysts. Novelty consisted in discovering a new functional policy and/or program in which models could play a part.

[2] A subsidiary, but equally useful, definition of 'policy' for our purpose is: 'a projected program consisting of desired objectives and the means to achieve them'.

[3] And computer-assisted information systems generally.

Vol. 7
1979

Distinctions which we have come to take for granted were first evolved during this period.

- Any policy or program issue to be dealt with at less than the national level became a 'subnational' one.
- 'Small-area' became the information domain of whatever geographically bounded area was the province a functional decision-maker (usually, but not always, *within* an urban area).
- 'Policy-making' became distinguished, although unclearly, from 'programming' and 'planning', with increased recognition that 'policy' had to do with what *ought* to be done (including what programming, planning, and budgeting needs to be done).[4]
- About the same time, the distinctions (originally made by Robert Anthony and Peter Drucker) between 'policy', 'planning', 'management', and 'operations' began to surface and became a part of the modelers' lexicon.
- 'Data' became increasingly differentiated from 'information', and the differential value of 'data' (and 'statistics') became a widely discussed topic.
- Correspondingly, information users began to surface, distinguishing themselves and their problems and needs, from data and information providers—and information 'middle-men' ('middle-people'?) began to emerge.

During these early years, new model-builders brought with them a concern for a number of new functional 'policy' areas—criminal justice systems, health manpower, regional development, ghetto redevelopment—as well as a variety of perspectives on the various dimensions of 'policy analysis'. Much light—and occasional heat—was generated in: clarifying the definition of urban and regional policy (i.e. the normative vs the programmatic, budgeting, and narrowly planning aspects); specifying what appeared to be necessary and sufficient conditions for the conversion of data into information for policy purposes (e.g. gaining access and working relationships with users and understanding the model of the world in which they work); and enumerating the problems of using models in policy-assisting activities (e.g. meshing the quantified with the subjective, meeting very difficult time schedules, and working interactively with the decision-maker).

Model-building was feeling its oats in these early years, but the best was yet to come.

The Optimistic–Romantic period. With the evolution of the Great Society program and the rapid increase in the use of hardware and software for urban analysis activities, model-building applications grew into the 1967–1968 period, which I characterize as the beginning of the 'optimistic-romantic' development phase. I dislike selecting an end-date for this attractive, 'fun' period, but I suspect it to be around 1970–1971. There were *so* many policy matters to decide, so many new uses of public powers and resources, so many new faces and ideas, and so many dollars. Data processing technology was coming into its own—at least for traditional data processing uses in the key urban administrative and management areas—finance, tax collection, personnel administration, payroll and other similar massive data processing related activities. That the characteristics of many of these successful data processing applications were seldom related to policy analysis, 'decision-making', or even physical and economic planning, was perceived by model-builders during these 'data-euphoric' late sixties. Many modelers had a planning and/or policy analysis background and a strong desire and an ambiguous technology for 'urban systems' to be of great assistance to urban and regional planners and policy-makers. The difficulties of these complex, data-transformation-to-information-for-decision-making urban roles were not lost upon these

[4] Some like to distinguish between 'policy' and each of three types of 'planning'—normative, strategic, and program—which are, in order, aimed at determining what *ought* to be done, *how* it can be done most effectively, and what *will* be done. 'Policy' and 'normative planning', which identifies preferred solutions to problems ('what ought to be done') are most often compared, since normative planning is directed toward 'goal setting'; however, there are also policy elements in strategic planning, defining how the normative goals can be achieved most easily, and 'program planning', defining the detailed means by which the programs (defined in the strategic policy phase) can be implemented.

early modelers—however, I believed they too readily projected what 'planners' *might* do with improved information. That planners were typically in a very complex environment, often unreceptive to what modelers might think of as totally 'rational', was part of the reason for policy-making resistance.

Not unexpectedly, there was a virtual avalanche of urban/regional models about the new planning, program analysis, budgeting and other 'futuristic' decision-making and policy-related decision-making applications 'using' urban/regional models. When reviewing this era, it is difficult not to wonder at the relative unsophistication—almost naiveté, at the 'world view' held by many of us. We believed in the 'goal articulation alternatives definition-alternatives evaluation and selection' view of the policy-making information users—whether or not they conformed more to the rationalistic model or the incrementalist (Lindblom–March–Simons) model of decision-making. The views about normative planning of John Dyckman, Frederic O'Reilly Hayes, Nathan Grundstein, Ira Robinson and Mel Webber (to mention a few) helped us connect policy, planning and data systems and showed us how 'information' *could* and *does* become power when used by public decision-makers. Information-rich analytic models were reported to the membership in almost all regional and urban functional areas (Laidlaw, Rogers, Teitz, Steger *et al.*).

Which came first—the euphoric optimism about method or the techniques to assist decision-making—is an academic question. During these years, we saw the birth and puberty of PBB, PPBRS, B–C, C–E, metropolitan model-building, social and urban indicators, and other quantitative and analytic abstractions for normative planning. Policy-makers and/or close aides—including big city and smaller town mayors, Congresspeople, Governors' offices and cabinet heads—reported to model-builder audiences on their uses of these methods. Many were anticipatory applications, in the throes of gestation. Much of the discussion focused on the impact of these decision-aiding systems—most of us presuming they would work as well as they were intended to, or better. As naïve as we were, modelers *did*, to their credit, believe these systems led to decision-assisting information and were *not* 'the decisions', *per se*.

Early seventies. A number of events and happenings took place as the effects of the change in administration after 1969–1970 became concrete:

1. Harder looks began to be taken at public programs—their efficiency, effectiveness, purposes. This emphasis—partly from the reigning philosophy that the best government was at best moderately sized—might have generated a sizable need for evaluative tools such as PPB, B-C, C-E, and other public policy assessment techniques. In turn, significant demand for urban and regional models generated subnational data. In *reality*, the assessment techniques were so embryonic that they proved not to be very useful and often *caused* more damage than good in dozens of overly literal applications. Policy analysis had *not* come of age—whether or not the first Nixon years really wanted it or not. The same lack of convincing utility, moreover, applied to other data consumption activities—metropolitan and regional planning, social indicators, community participation, and a host of urban development programs. Clearly many applications were at least partially successful; however, in some cases, there was a good deal of *ad hoc* need for additional data collection efforts to supplement the standard data bases (e.g. for social experiment purposes).

2. At the same time, faith and dollars in public programming decreased. Partly attributable to Vietnam, partly to the plethora of new programs which had to be weeded out: the result was that analytic/modeling projects were often caught in the middle. In promising new agencies the resultant decline in agency power carried with it, *pari passu*, the passing of many new and promising policy-making systems opportunities and careers.

3. A number of problems offered substantial impediments to expanded modeling and urban data systems use, for policy-making, in this period: confidentiality, security, and privacy issues: proprietary data; standardization; and system compatibility require-

ments. These were *not* new issues or problems, but they attained a head of steam in these years—right- and left-wing organizations and public interest groups finding a common interest in specific anti-analytic activities.

4. The true, total *costs* of urban modeling became more apparent, during this period, than the true *benefits*. While intensive assessments of the use of such models were still in the embryonic stages, only limited *benefits* from the use of models in the policy process were showing up. Assessing modeling utility is bound to be a difficult matter—and, as is usual in benefit–cost assessments, we tend to know how to measure costs, first. The tendency is a conservative one, at best, and serves to act as a damper on the enthusiasm for most modern, problem solving technologies. Some, such as computer hardware, offer such immediate cost-reduction opportunities to traditional, repetitive functions, that their benefits are more obvious. Benefits are considerably less obvious in modeling/policy-assisting roles.

5. A final 'damper' in this period was the relatively severe attack on analytic, policy analysis models (Brewer, Lee, Shubik, Wildavsky and others). Since many of the analytic aids to policy analysis were tied to models generated from requirement analyses of decision-making processes, the intellectual efforts of model builders became directed to response and rationalization, rather than aggressively moving into new and challenging policy-aiding assignments.

For these and other reasons, the early to middle seventies saw a lowering of expectations for urban models in policy-assisting roles, and an increased understanding of what role and responsibility such models *can* and *could* play in normative public sector decision-making.

The Now period. To some extent, we have come full-circle since 1963. Model-builders are significantly more aware of the difficulties of decision-making. We are well aware—as is exemplified by Kraemer's work [10] as well as in numerous ORSA/TIMS, URISA and other papers in 1975 through 1977 and in the public policy literature generally [2, 3, 6, 8, 10], that we are indeed dealing with what Russell L Ackoff has called the "messy management— of wicked problems—buzzing with myriad confusions" [1].

In fact, it is the knowledge of the complexity of the policy analysis process which I feel gives the current model-builder renewed faith in his/her abilities to assist policy-makers. The relative success of a number of recent social experiments to help in framing human resource and housing policies; the increased visibility of information systems' operations at the side of several big city (and regional) chief operating officials, in at least a dozen major metropolitan areas; the increased sophistication of information system evaluation tools to 'tease out' the benefits of models in policy-making; renewed use of PPB (and its modified zero-based-budgeting form) and social indicators in several urban areas; the presence of lowered but more realistic expectations; increased understanding and communication about system approaches of past and present big city mayors; incremental solutions to the problems of confidentiality, security and privacy; increased attention to user feedback in policy analysis; greater focus on transferability, at the policy level, of analytic methods and accompanying data bases; increased attention to making outputs of analytic models useful to policy-makers; and generally more sensitivity and accommodation to policy analysts' special needs (e.g., for speedy and well-represented information): these are some of the reasons for resurgence of interest in policy-assisting models.

WHERE TO NOW?

I believe the 360° movement since 1963 has been a useful, fruitful and adventurous intellectual voyage. I am delighted to have been a part of it, to have experienced it with many members of the modeling profession, and to share with my colleagues the renewed faith. Less romantic and optimistic, perhaps, but realistically:[5]

1. We have learned to disabuse ourselves of the notion that it is simple to make analytic system inputs to policies.

[5] I thank my colleague William C. Garrison for his suggestions, here.

2. We have learned a good deal about 'what it takes' to make inputs to policies—increased understanding of decision-making, early information system development, and user feedback being some of these lessons.

3. There are large time constraints that have to be accepted, patiently, by model-building policy-assisting providers. We have taken fifteen years to get this far, I think we can take several years to better understand from where we have come and to where we are going.

Most importantly, perhaps, is that we need to increase our emphasis on learning more about which models, in what contexts, do well, and which don't and why. My colleagues and I, for example, have been retained by regional planning agencies to review their batteries of models, for future applications and problems facing their urbanized areas [4].

Our objective cannot be so grandiose as to determine which analytic techniques are most contributive of effective planning because—as correctly pointed out by Peter House (Letters to the Editor, *JAIP*, January, 1974)—model-building's inadequacies should not be held responsible for what 'planning' can and cannot do and is and is not doing. The use of modeling techniques arises, more or less, as a 'derived demand' from what planners (think they) want to do and hope to achieve. Therefore, the onus for those who seek to develop performance measures for modeling approaches lies in working with the derived demand—not the ultimate solution to planning issues, *per se*. As such, modeling-yielded information—viewed in the broadest sense as any resource which planners find useful in achieving their objectives, including 'educational' purposes—must be seen in perspective along with other similar resources. Modeling-produced information, including levels of understanding produced in developing a modeling approach, needs to be assessed, in cost/effectiveness terms where possible, along with other methods for producing similar information. Nothing is absolute, here: the quality, credibility and 'validity' of all such information needs to be assessed, relative to each of the methods capable of producing such information. The planner, as both the broker of such information as well as its direct recipient, is in the best position to assess what different 'levels' and 'degrees' of quality, credibility and validity will mean to his various clients. In the production of his planning products and the improvement of his planning process, each planner responsible for making such choices must deal not with modeler rhetoric, or those who review past efforts, but rather with *his* needs, *his* clients, and alternative methods available to *him*.

I cannot spell out the performance measures, here; a more comprehensive effort, perhaps involving an evaluative process [16] needs to be set in motion. I believe Gary Brewer, Douglas Lee and others have made their greatest contribution in awakening the planning and analysis professions to the need for evaluation, for self-assessment, for feedback to ongoing efforts. Personally, I feel that most performance measures to date have been 'input' oriented—and, perhaps, that is all that can be accomplished, now. But no one can be fully satisfied with 'unreasonable rigor' (one of Lee's suggested criteria) because, again, this is a 'motherhood' objective, unless it is compared with the use of such information—the errors of commission and omission that would arise with and without it, and the societal and private costs (and the probabilities associated with such costs) attached to these various types of errors.[6]

Perhaps the most useful insights the recent assessments have produced are those related to the dynamics of acceptance. The entire sociology of technological innovation and diffusion could well be applied to this problem—including the initial total rejection of 'foreign' models, to partial rejection, and then—in some cases—acceptance. Quite obviously, fairly lengthy time periods may be involved in assessing how a particular use or user is affected by a specific technique. Or how a particular technique initiated in one place, for one purpose, is used in other places, for similar or even different purposes. The current emphasis on transferability of methodology, in specific Federally spon-

[6] Development of such criteria merit research: William Alonso's "Choosing and Building Models for Prediction", the Bureau of Public Road's *Systematic Sensitivity Analysis of the Urban Travel Forecasting Process*, and JW Tukey's "Propagation of Errors, Fluctuations and Tolerances" are all good beginnings in the search for measures of informational benefit, but a good deal of further investigation and application is needed along like lines.

sored analytic projects, can both learn from the earlier efforts, and can help increase the effectiveness and efficiency of current efforts. The Brewer, Lee *et al.* observations on the differential attitudes and motivations of the users, developers, politicians, and the general public, provide helpful insights into the breadth of perspective needed to produce an effective methodology. But they also point up the substantial length of time needed to make judgements about the acceptability and diffusion of a technique: a city may be better off because it has been through an unsuccessful attempt to apply a model—perhaps it will be more or less willing to accept the next innovation; or the unsuccessful application in City 1 may point to how City 2 should successfully apply the method, etc. Future evaluations need to consider the *sociology* of the acceptance process.

Personally, I would like to see considerable attention paid to positive approaches to model-building evaluation, assessment, and feedback. Some possible ideas include the development and use of:

- Peer Review Groups,
- User Group Reviews,
- Central repositories for application's data e.g. methods, costs etc.

I look forward, perhaps in 1989, to take 'more stock of the past'—and be able to comment most positively on the favorable trends I feel we have begun to experience in the latter half of this decade.

REFERENCES

1. ACKOFF R (1977) The corporate rain dance. *The Wharton Magazine* **1,** 36–41.
2. ANGRIST SS, BELKIN J & WALLACE WA (1976) Social indicators and urban policy analysis. *Socioecon. Plann. Sci.* **10,** 193–198
3. BAKER FR, MICHAELS RM & PRESTON ES (1975) *Public Policy Development Linking the Technical and Political Processes,* John Wiley, New York, USA.
4. CONSAD RESEARCH CORPORATION (1978) *Review of Puget Sound Council of Governments Computerized Planning Models and Data Sources.* Prepared for the Puget Sound Council of Governments.
5. DROR Y (1970) Prolegomena to policy sciences. *Policy Sci.* **I,** 135–150.
6. FRANKEL RJ & FRANKEL BF (1969) Microenvironmental responses to changes in the urban environment. In *Quality of the Urban Environment* (PERLOFF HS ed), Johns Hopkins Press, Baltimore, USA.
7. HARRIS B (1967) Paper prepared for the Committee on Urban Economics: analytical and policy issues.
8. JOHNSON N & WARD E (1972) Citizen information systems: using technology to extend the dialogue between citizens and their government. *Mgmt Sci.* **19,** P-21–P-34.
9. KILLBRIDGE M (1967) *Urban Planning Models.* Graduate School of Business Administration, Harvard University, USA.
10. KRAEMER KL *et al.* (1976) A future cities survey research design for policy analysis. *Socio-Econ. Plann. Sci.* **10,** 199–211.
11. LAKSHMANAN TR & STEGER WA (1967) *Plan evaluation methodologies: some aspects of decision requirements and analytical response.* Paper prepared for the Highway Research Board Conference on Urban Development Models, Dartmouth College, *Urban Development Models,* HRB Special Report **97.**
12. LAMB DD (1967) Research on existing land use models. Southwestern Pennsylvania Regional Planning Commission.
13. LEWIN AY & SHAKUM MF (1976) Situational normativism: a descriptive normative approach to decision-making and policy sciences. *Policy Sci.* **III,** 1–10.
14. LOWRY IS (1967) *Seven models of urban development! a structural comparison.* RAND Paper P-**3673.**
15. PITTS FR (1964) *Scale and purpose in urban simulation models.* Discussion paper prepared for Conference of Strategy for Regional Growth.
16. STEGER WA (1974) Letter to the editor. *Jl Am. Inst. Planners.* **31,** 166–172.
17. STEGER WA (1977) Information systems inputs to policies: a bench-mark summary, 1963–1977. *Pap Fifteenth An. Conf. Urban & Reg Inf. Systems Assoc.* **I,** 30–39.
18. STEGER WA (1966) The realities of simulation exercises for city planning. Paper delivered before the New York City Chapter of the American Statistical Association, *Proceedings on Simulation in Business and Public Health.*
19. STEGER WA (1965) Review of analytic techniques for the CRP. *Jl Am. Inst. Planners,* **31,** 166–172.
20. STEGER WA (1967) Urban simulation: talking stock of the recent past. Prepared for the Center for Regional Studies, University of Kansas, USA.
21. TRAFFIC RESEARCH CORPORATION (1963) Review of existing land use forecasting techniques. Boston Regional Planning Project, Massachusetts Transportation Commission

ADDRESS FOR CORRESPONDENCE: *Dr Wilbur A Steger, CONSAD Research Corporation,* 121 *North Highland Avenue, Pittsburgh, Pennsylvania* 15206, *USA.*

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 553 to 556
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0553$02.00/0

The Lognormal Approximation to the Lead Time Demand in Inventory Control

PANDU R TADIKAMALLA

University of Pittsburgh

(*Received February 1979; in revised form April 1979*)

This paper is intended to show that the lognormal distribution can be used to approximate the lead time demand in inventory control problems. The computations involved with this approximation are shown to be simpler compared to the gamma distribution approximation. Some pertinent properties of the lognormal distribution are summarized. Numerical examples are given in two cases of distributed demands and lead times and in each case the lognormal approximation results are compared with the exact and/or other approximations.

INTRODUCTION

TWO OF THE central problems in inventory control are (i) to determine the protection level (the probability of not having a stock out) for a given reorder level (see, for example, [4, 6]) or (ii) to determine the reorder level so as to have a given protection of not having a stock out (see [14]). These quantities will depend upon the distribution of the lead time demand, $f(x)$, given by

$$f(x) = \int_0^\infty f(x|L)\, g(L)\, dL \qquad (1)$$

where $g(L)$ is the distribution of the lead time (assuming $0 \leqslant L < \infty$) and $f(x|L)$ is the distribution of the demand for any given L.

Although a variety of distributions can be used to represent the lead times ($g(L)$) and the demands ($f(x|L)$), rarely does there exist a combination of these distributions that would yield a familiar distribution for the lead time demand $f(x)$. Hadley and Whitin [9] showed that $f(x)$ is a negative binomial if $g(L)$ is gamma and $f(x|L)$ is Poisson. Often the normal and the gamma distributions are used to represent the demands and/or the lead times (see [4, 14]). However, in these cases the resulting distributions of the lead time demand are too complex, [4] and some approximations to these distributions are sought. The gamma distribution (see [5]) and the Weibull distribution (see [18]) have been advocated as approximations to the lead time demand. Although more general and versatile families of distributions such as Pearson system [16], Johnson system [11], Burr distribution [7] and the generalized lambda distributions [8] can be used as models to represent the lead time demands, their use may be limited because of the complexity in parameter estimation and/or further computations of the protection and reorder levels and other quantities.

In this paper we advocate that a two-parameter lognormal distribution can be used as an approximation to the lead time demand and the approximation is shown to be accurate for two cases. Since the percentiles and cumulative distribution function values of the lognormal distribution can easily be calculated from the corresponding standard normal values (which are extensively tabulated and are reproduced in almost all statistics books), the re-order levels and the protection levels can easily be obtained. Also the parameters of the lognormal distribution can easily be obtained without the use of tables or numerical procedures using the method of moments or the maximum likelihood methods.

PROPERTIES OF THE LOGNORMAL DISTRIBUTION

The probability density function (*pdf*) of the lognormal distribution can be given in many

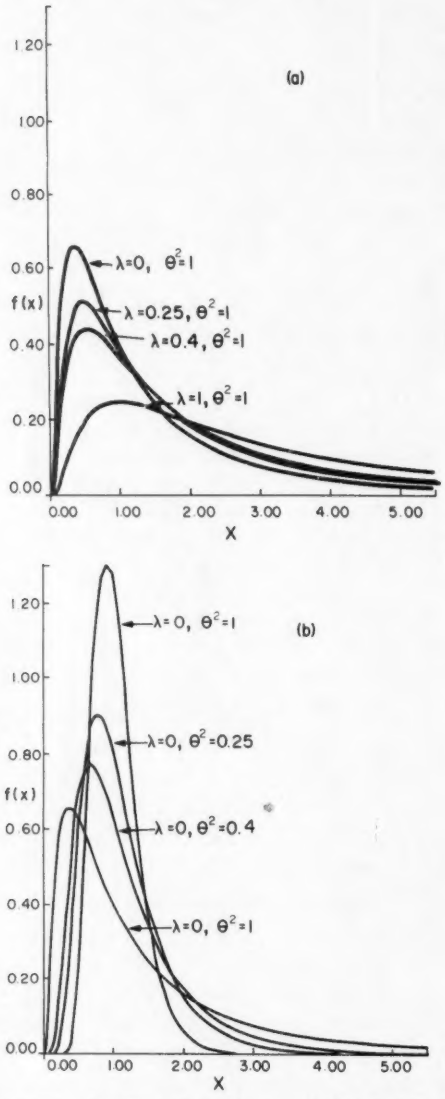


FIG. 1. *The densities of the lognormal distribution for selected values of* λ *and* θ^2

different ways. We define that the variate X has the lognormal distribution if the variate $Y = \ln X$ has the normal distribution with mean λ and variance θ^2. Also note that the variate $Z = (Y - \lambda)/\theta$ has the standard normal distribution given by $h(z) = \exp(-z^2/2)/\sqrt{2\pi}$.

(1) The *pdf* of the lognormal variate is given by

$$f(x) = \exp\{-(\ln x - \lambda)^2/2\theta^2\}/(\theta x\sqrt{2\pi}),$$
$$x > 0,\ \theta > 0,\ -\infty < \lambda < \infty. \quad (2)$$

θ is called the shape parameter and, λ is the scale parameter.

(2) The k^{th} moment of lognormal variate, X, is given as

$$E(X^k) = \exp\{k\lambda + k^2\theta^2/2\}. \quad (3)$$

In particular the mean μ and variance σ^2 are

$$\mu = \exp\{\lambda + \theta^2/2\} \quad (4)$$

and

$$\sigma^2 = \exp(2\lambda + \theta^2)\{\exp(\theta^2) - 1\} \quad (5)$$

(3) The lognormal distribution is positively skewed (the skewness depends only on θ) and as $\theta \to 0$, (2) $\to$ normal distribution.

(4) The lognormal distribution is unimodal with mode

$$x = \exp(\lambda - \theta^2).$$

(5) For X in (2), $\mu >$ median $>$ mode and the ratios μ/σ and μ/mode are functions of only the shape parameter θ,

$$\mu/\sigma = \{\exp(\theta)^2 - 1\}^{-1/2} \quad (6)$$

and

$$\mu/\text{mode} = \exp\{3\theta^2/2\}. \quad (7)$$

The densities of (2) for selected values of θ^2 and λ are shown in Fig. 1.

FITTING THE LOGNORMAL DISTRIBUTION TO THE LEAD TIME DEMAND

The estimation of the parameters λ and θ presents very few difficulties. A detailed account of various estimation methods can be found in [2, 3, 12]. Briefly:

(1) Let the mean and variance of the lead time distribution be L_1 and L_2 and those of the demand distribution per unit lead time be S_1 and S_2. It is known that the mean and variance of the lead time demand distribution D_1, D_2 are

$$D_1 = L_1 S_1 \quad (8)$$

and

$$D_2 = L_1 S_2 + L_2 S_1^2$$

The parameters λ and θ of the lognormal distribution (2) can easily be determined, to give the mean D_1 and variance D_2, as follows:

(a) From (6), solving for θ, we obtain

$$\theta = \{\ln(1 + D_2/D_1^2)\}^{1/2}. \quad (9)$$

(b) Having obtained the value of θ, the value of λ can be estimated from (4);

$$\lambda = \ln D_1 - \theta^2/2. \quad (10)$$

Vol. 7
1979

(2) If the actual lead time demand data, say x_1, $x_2, \ldots\ldots x_n$, is available the maximum likelihood estimators of λ and θ can be easily obtained.

If $z_i = \log x_i$, then

$$\lambda = \bar{z} = \sum_{i=1}^{n} z_i/n$$

and

$$\theta = \left\{ \sum_{i=1}^{n} (z_i - \bar{z})^2/n \right\}^{1/2} \tag{11}$$

Note that these quantities are simply the mean and variance of the logarithms of the observed data.

(3) If only estimates of the mean demand (D_1) and the most likely demand (mode, M) during the lead time can be made, the values of λ and θ can be estimated as follows:

(a) from (7) we obtain

$$\theta^2 = (2/3)\ln(D_1/M) \tag{12}$$

(b) After θ is estimated the value λ can be obtained using (10). Note that no tables or numerical procedures are required for estimating the parameters of the lognormal distribution by any of these methods.

PROBABILITY OF STOCK OUT

In the following sections we take the distribution of the lead time demand to be the lognormal distribution given by (2). Under a classical reorder level inventory system, the protection, (probability of not being stock out) P_R,with a reorder level R, is

$$P_R = \int_0^R f(x)\,dx = F(R) \tag{13}$$

where $F(.)$ is the *cdf* relating to the *pdf* $f(.)$. With the help of the standard normal tables, $F(R)$ for (2) can easily be obtained. Let $z = (\ln R - \lambda)/\theta$. The required probability (protection) is simply the *cdf* value of the standard normal corresponding to z. Thus with the lognormal approximation no numerical procedures are required to obtain the protection level. In fact the standard normal table look up operation can be also avoided if a simple and reasonably accurate approximation (see [10, 15, 17]) is used for the normal distribution. Also most modern scientific calculators do have the capability of calculating the *cdf* and the inverse of the *cdf* of the standard normal distribution.

Example 1: Gamma distributed lead time and normal distributed demand (see [4]) Consider the case $S_1 = 5$, $S_2 = 1$, $L_1 = 4$ and $L_2 = 1$ so that, from (8), $D_1 = 20$ and $D_2 = 29$. The parameters of the lognormal distribution θ and λ can be obtained from (9) and (10), respectively; $\theta = \{\ln(1 + 29/400)\}^{1/2} = 0.26456$ and $\lambda = \ln 20 - 0.26456^2/2 = 2.96073$. The protection level P_R for a reorder level of $R = 16.1555$ $(= 3\sqrt{D_2})$ can be obtained as follows: Let $z = [\ln R - \lambda]/\theta = -0.67459$. From the standard normal tables (see [1]) the probability (*cdf*) corresponding to $z = -0.68$ is given by $1 - 0.75174 = 0.24826$. For this example the exact value of $P_R = 0.24917$, Burgin's approximation value is 0.25010, and Tadikamalla's Weibull approximation value is 0.23905 (See [18]).

DETERMINATION OF REORDER LEVEL R

As in the case of Moonen [14], suppose we are interested in determining the reorder level R for a given protection level $P_R = p$. We want to find R for a given p in the following equation:

$$p = \int_0^R f(x)\,dx = F(R).$$

Therefore

$$R = F^{-1}(p). \tag{14}$$

The value of R in (14) can easily be obtained in the case of the lognormal approximation. Let z_p be the standard normal 100pth percentile. Then

$$R = \exp(\lambda + z_p\theta). \tag{15}$$

Again no numerical or iterative procedures are required to obtain the reorder level R.

Example 2: Normal distributed lead time and gamma distributed demand. For easy comparison, let us consider the same example given in [18]:

$L_1 = 10$, $L_2 = 4$, $S_1 = 6$ and $S_2 = 3.6$. Then $D_1 = 60$ and $D_2 = 180$. Again from (9) and (10) the values of θ and λ, respectively, can be obtained as $\theta = 0.22088$ and $\lambda = 4.06995$. We want to obtain the reorder level R, say, for a protection level of 90% i.e. $P_R = p = 0.9$. The standard normal percentile corresponding to $p = 0.9$ is $z_p = 1.28155$. Then from (15), the reorder level $R = \exp\{4.06995 + 1.28155 \times 0.22088\} = 77.712$.

TABLE 1. COMPARISON OF COMPUTATIONAL COMPLEXITY FOR THE GAMMA, WEIBULL AND LOGNORMAL DISTRIBUTIONS

	Requirements				
	Parameter estimation		Inventory calculations		
Distribution	Moments	MLE	Protection level	Reorder level	Potential lost sales
Gamma	None	T and/or NP	T and/or NP	T and/or NP	T and/or NP
Weibull	T	T and/or NP	None	None	T and/or NP
Lognormal	None	None	None*	None*	None*

T = Special tables, NP—Numerical procedures.
* Requires standard normal tables.

According to Moonen this value is 77.4, and Tadikamalla's Weibull approximation value is 76.87. Once again no tables or numerical procedures are required for obtaining the reorder level.

SUMMARY AND COMMENTS

It has been shown that the lognormal distribution can very well be used to approximate the lead time demand in certain cases of distributed demand and lead times. The reasons for the potential success of the lognormal distributions in these inventory control problems are:

(i) The lognormal distribution can represent a wide variety of skewed frequency curves.

(ii) The *cdf* and the percentiles (inverse of the *cdf*) of the lognormal distribution can easily be obtained from the standard normal tables, thus making the calculation of protection levels and reorder levels easier.

(iii) No tables or numerical procedures are required for estimating the parameters.

It is interesting to note that all the three distributions, the gamma, the Weibull and the lognormal 'look' very much alike when their corresponding means and variances are equal (see [13]). All these three distributions seem to provide accurate enough approximations to the lead time demand in the particular cases considered. The computational complexity (or simplicity) involved in using these distributions for these applications varies considerably ranging from no tables, to table look up, to numerical procedures. (see the Table 1).

REFERENCES

1. ABRAMOWITZ M. & STEGUN IA (1972) *Handbook of Mathematical Functions.* US Government Printing Office, Washington, DC, USA
2. AITCHISON J & BROWN JAC (1957) *The Lognormal Distribution*, Monograph **5**, Cambridge University Press, London, UK.
3. BROWN RG (1959) *Statistical Forecasting for Inventory Control*, Appendix C—The lognormal distribution, McGraw-Hill, New York, USA.
4. BURGIN TA (1972) Inventory control with normal demand and gamma lead times. *Opl Res. Q.* **23**, 73–80.
5. BURGIN TA (1975) The gamma distribution and inventory control. *Opl Res. Q.* **26**, 507–525.
6. BURGIN TA & WILD AR (1967) Stock control-experience and usable theory. *Opl Res. Q.* **18**, 35–52.
7. BURR IW (1973) Parameters for a general system of distribution to match a grid of α_3 and α_4. *Commun. Stat.* **2**, 1–21.
8. DUDEWICZ EJ, RAMBERG JS & TADIKAMALLA PR (1974) A distribution for data fitting and simulation. *ASQC Tech. Conf. Trans.* **28**, 407–418.
9. HADLEY G & WHITIN TM (1963) *Analysis of Inventory Systems.* Prentice-Hall, Englewood Cliffs, New Jersey, USA.
10. HASTING C (1955) *Approximations for Digital Computers.* Princeton University Press, Princeton, USA.
11. JOHNSON NL (1949) Systems of frequency curves generated by methods of translation, *Biometrika* **36**, 149–176.
12. JOHNSON NL & KOTZ SS (1970) *Continuous Univariate Distribution I.* Houghton Mifflin, New York, USA.
13. LITTSCHWAGER JM (1965) Results of a gamma, lognormal, and Weibull sampling experiment. *Industrial Quality Control* **XXII**, 124–127.
14. MOONEN HJM (1962) The determination of re-order levels when demand and delivery times have gamma and normal distributions, respectively. *Statistica Neerlandica* **16** 113–120. (original in Dutch).
15. PAGE E (1977) Approximations to the cumulative normal functions and its inverse for use on a pocket calculator. *Appl. Statist.* **26**, 75–76.
16. PEARSON K (1895) Contributions to the mathematical theory of evolution II. Skew variations in homogeneous material. *Phil. Trans. R. Soc. Lond., Ser. A* **186**, 343–414.
17. RAMBERG JS & SCHMEISER BW (1972) An approximate method for generating symmetric random variables. *Comm. Ass. Comput. Mach.* **15**, 987–990.
18. TADIKAMALLA PR (1978) Applications of the Weibull distribution in inventory control. *Jl Opl Res. Soc.* **29**, 77–83.

ADDRESS FOR CORRESPONDENCE: *Professor Pandu R Tadikamalla, Graduate, School of Business, University of Pittsburgh, PA 15260, USA.*

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 557 to 561
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0557$02.00/0

Dynamic Aspects of Line Balancing

DOMINIC SCULLI

University of Hong Kong

(*Received August 1978; in revised form March 1979*)

The article explores the dynamic nature of line balancing. Models are developed to help management with the problem of manning a line to produce a specified output. The models quantify the fluctuations in both excess inventories and the number of operators required.

INTRODUCTION

MOST LITERATURE and theory on line balancing is concerned with the initial assignment of work elements to stations. A great deal of research effort has been directed towards balancing the line before it is constructed [1, 6–8]. However, the static balancing problem is rarely solved in such a way that the line operates at maximum efficiency. Technological restrictions, and these may include the nature of the work and the type of equipment required for the operations, mean that it is often impossible to assign work to stations so that the desired output is achieved with 100% line efficiency. Fluctuations in standard rates and equipment failure will also contribute to the imbalance. This will manifest itself as idle operators and shortage or excess of work in progress inventory and finished goods. Idle operators will probably attempt to appear busy by adjusting their output rate to that of the previous station. The excess inventory cannot be allowed to go on growing or be ignored so the line must be adjusted, operators moved to bring it under control. The nature of a production system is in fact dynamic and the balancing problem still exists even if an optimum solution to the problem has been obtained in its static state.

The dynamic balancing problem does not appear to have attracted as much interest as the static problem. Davis [2, 5] looked at the problem of reducing excess work in progress inventories to a desired level. He deals with a situation where the inventories have gone out of control and reduces them to a desired (management decision) level by assigning fewer operators and fractions of an operator. Later [3], he proposed a heuristic procedure for determining this level. In all this work mobility of operators is assumed at any time during a shift. However, some lines will not have the degree of operator mobility required to apply this theory. Full mobility requires equal levels of training for all tasks quite apart from willing cooperation of operators. It is common practice to assign line operators at the beginning of a work period and then to run it under constant manning. Davis and Simmon [4] considered the work period constant manning, and a shift by shift line manning schedule is computed. However, in some situations it would be very difficult to adhere to the schedule, especially when there is variability in operator standard rates. This variability is partly responsible for the imbalance and consequently the excess work in progress inventories, and also for the size of variations in operators required.

The hypothesis of this paper is that the dynamic balancing problem has to be divided into two parts. First, determination of the average in-process inventories and variations in the number of operators required to run the line at a given output. Second, the shift by shift allocation of operators, which is predominantly determined by the operators available and the in-process inventory levels at the start of the shift. The model developed in this paper is an attempt to solve the first problem. The exten-

sion of this model to help with second problem is discussed in the conclusion.

NOTATION AND ASSUMPTIONS

α is the probability that the output from any station will be sufficient to keep the direct successor station working. In the case of the final station it is the protection level against finished goods stock-outs.

r_i is the operator output rate at station i in units per unit time ($i = 1, 2, \ldots m$).

p_i is the coefficient of variation of the operator output rate at station i. It is assumed that the operator output rate has a normal distribution with mean r_i and variance $p_i^2 r_i^2$.

T is the time, work session, during which the line will run unadjusted (constant manning).

X_i is the excess work in progress at station i, i.e. waiting to be processed by station i.

n_i is the number of operators assigned to station i.

E is the excess finished goods stock.

R is the constant output rate (demand rate) from final stock.

$S(\alpha)$ is the inverse cumulative density function of the unit normal distribution evaluated at α.

MODEL AND OBJECTIVES

The long-term objective in assigning operators to the final station m is to meet demand at the specified protection level α. For any work period T we would like:

$$Tn_m r_m = TR - E + S(\alpha) p_m r_m \sqrt{Tn_m} \tag{1}$$

For intermediate stations the aim is to keep the direct successor station working during the period with the given probability α. So ideally in any period we should have:

$$Tn_i r_i + X_{i+1} - Tn_{i+1} r_{i+1} = S(\alpha) \sqrt{T(n_i r_i^2 p_i^2 + n_{i+1} r_{i+1}^2 p_{i+1}^2)} \,. \tag{2}$$

The above equations will give an assignment for the existing excess inventory figures. However, the number of operators assigned, n_i, will need to be rounded to the next highest integer. The expected excess work in progress inventories at the end of the period X_i^* will be:

$$X_i^* = Tn_{i-1} r_{i-1} + X_i - Tn_i r_i.$$

The expected excess finished goods inventory E^* will be:

$$E^* = Tn_m r_m + E - TR$$

The new excess inventory figures can be used to assign operators in the next and subsequent work periods. Repeated application will show the variations in operators required to run the line, and also the average work in progress inventories generated.

EXCESS INVENTORIES

The excess inventories generated can be divided into two components. The first what may be called the base inventory, is generated by the difference in station rates. If we assume no variability in station rates and perfect information on excess inventories, the base inventories can be determined as follows. If we try to maintain an inventory close to zero, and in each work period we can only increase it by an amount a units or reduce it by b units, then the average inventory will be $(a + b)/2$ provided a is not divisible by b or b divisible by a. It will be $a/2$ if b divides a', or $b/2$ if a divides b. At each station we can control work in progress inventories through an additional operator, increase or reduce the output by Tr_i. If w is the additional output required to keep the next station working, then excess work in progress ahead of station i can be increased by $Tr_i - w$ or reduced by w. Hence the average work in progress inventory will be $Tr_i/2$. If the operator rate of the successor station r_{i+1} is divisible by the station rate, r_i no work in progress should be necessary. However, this is an ideal situation easily upset by a slight disturbance. The total base in inventory will be:

$$\sum_{i=2}^{m+1} Tr_i/2$$

$$r_{m+1} = R.$$

If r_{i-1} divides r_i, we exclude the term from the summation.

The other component of the excess inventories will be generated by variations in operator rates and the specified protection level α. The mean number of operators assigned at any station will be R/r_i. The difference between the

Vol. 7
1979

output of station i and the input to station $i+1$ will have a normal distribution with mean zero and variance equal to the expression under the square root sign in (2). The specified protection level α implies buffer inventories equal to:

$$\sum_{i=1}^{m} S(\alpha)\sqrt{TR(r_i p_i^2 + r_{i+1} p_{i+1}^2)}$$

$$r_{m+1} = R$$

$$p_{m+1} = 0.$$

The above shows that excess inventories can be reduced by making the work period small, remembering that this is the time frame during which we want the line to run unadjusted. An obvious conclusion, since frequent adjustments, reallocation of operators, gives better control over excess inventories. However, operators may not tolerate or be adequately trained for frequent station changes.

OPERATORS REQUIRED

If the standard rates are assumed to be constant, the variation in operators required at the various stations is due to the difference between station rates and the output rate R. The maximum and minimum operators required may be computed as follows:

$$n_i = n_{i+1}\, r_{i+1}/r_i$$
$$n_m = R/r_i. \qquad (3)$$

If the divisions are rounded to the next lowest integer, we get the minimum; rounding to the next highest integer will give the maximum. If r_i divides r_{i+1}, rounding is not necessary.

When variability in rates is introduced, the theoretical maximum is infinite. However, in this situation it is more constructive to look for the α confidence limits on the maximum number of operators required. If an α protection level is used against finished goods stock-outs, the upper α limit on the operators required at the final station can be computed using equation (1) with $E = 0$. Equation (3) can be used for the other stations since the system is already protected against finished goods stock-outs and an average work in progress inventory can be assumed to exist between intermediate station.

These operator variations do not necessarily imply idle operators. Almost all production lines produce some substandard output which requires rework and dismantling. A supply of this work can be made available at each station, so that operators can be withdrawn from the line when necessary and asked to work on it.

EXAMPLE PROBLEMS

The first example, see Table 1, is a four-element single-stage line. This type of line is heavily constrained by technological factors and there are large differences in station rates. The inventories and operators required for the specified output are shown in Tables 2 and 3.

TABLE 1. EXAMPLE PROBLEM 1

Station	Number (i)	Mean standard time units per hour	Standard time variance $p_i = 0.2$
Assemble	1	5.1	1.04
Bond	2	8.2	2.69
Test	3	10.7	4.58
Pack	4	6.3	1.59

The average output from finished stock is constant at 50 units per hour, and the work period is 8 hours. It is also assumed that the line is in a start-up position, i.e. no excess finished goods or work in process inventories. The protection level against stock-outs α is 99%.

TABLE 2. AVERAGE WORK IN PROGRESS AND FINISHED GOODS INVENTORIES

Station no. i	Base imblance results for 300 work periods	Theoretical results			Results 1000 work periods	Results of a 1000 work periods simulation
		Base imbalance	Variation in rates	Total		
2	20.30	20.40	33.98	54.38	53.20	54.40
3	34.40	32.80	40.52	73.32	72.70	71.90
4	42.60	42.80	38.43	81.23	79.30	78.30
Finished goods	25.40	25.20	23.39	48.59	47.50	48.20
Total	122.70	121.20	136.32	257.52	252.70	252.80

$R = 500$, $T = 8$, $P_i = 0.2$ for all i, $\alpha = 0.99$.

BLANK
PAGE

Vol. 7
1979

MEMORANDA

ol. 7
979

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 565 to 575
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0565$02.00/0

Quantitative Models for Optimal Rest Period Scheduling: a Note

INTRODUCTION

IN A RECENT paper, Gentzler, Khalil and Sivazlian [2] developed optimal policies for rest period scheduling for two deterministic models intended to be representative of decrements in work output as a result of fatigue and other factors. The two models were:

(i) $v(t) = b - at$
(ii) $v(t) = e^{-at}$

where

$v(t)$ = the instantaneous rate of work or the work rate function which starts at $t = 0$ for each work period,
a = the decay rate for either model, and
b = the maximum performance rate for the linear model.

Optimal policies were developed for an infinite time horizon under the assumption that some constant rest period of sufficient magnitude would permit a return to the maximum output rate. Under these conditions, optimal policies consist of alternating periods of work (with constant length x^*) and rest (with constant duration θ).

Gentzler *et al.* [2] examined the x^* policy via simulation for one parameter set for each of the two models with a finite horizon which did not correspond to an integer number of cycles for x^*, in part, to "... serve as a sensitivity analysis of the effects of different policies on work output..." [2, p. 218]. The infinite horizon policy of x^* was compared in the simulation with that of from zero to five evenly spaced rest periods. Evenly spaced meant that all work periods were of equal length with a complete work period at both the beginning and the end of the time horizon. The results of the simulation indicated that:

> the x^* policy was at least as good as any other, even though there was not an integer number of cycles. However, the policies for two and three evenly spaced rests resulted in total work output as large as the x* policy. This indicates some insensitivity of the work output for rest policies for these cases. [2, p. 219]

The sensitivity of work output to various rest policies could have been investigated analytically since the two models were deterministic. In addition, more general conclusions concerning sensitivity could have been drawn if more than one specific parameter set for each of the two models had been investigated.

This paper reports the results of an analytical policy comparison and suggests a simple procedure, in lieu of dynamic programming, for the determination of optimal finite horizon policies.

ANALYTICAL SENSITIVITY ANALYSIS

An analytical policy comparison was conducted for both models using all possible combinations of the following parameter values:

$a = 1/10, 1/20, 1/30, \ldots, 1/400$
$\theta = 1, 4, 7, 10, 13, 16, 19, 22$ (min)
$b = 1.0$
$T = 240.0$ (min)

where T is the length of the finite time horizon. Hence, each of the two models was investigated using 320 different parameter sets. The results confirmed that, in the neighborhood of the optimal policy, the work output of both models was rather insensitive with respect to the number of rest breaks. However, a rest break policy which is not in the neighborhood of the optimal policy may substantially limit work output. For example, the optimal work output of the exponential model (with the parameter set $a = 0.10$, $\theta = 7$ min and $T = 240$ min) is 359% higher than the work output with one rest break. In general, for either model, this type of result becomes more severe as the decay rate increases and as θ decreases.

As would be expected, the analytical analysis also confirmed that the optimal policy for a finite time horizon always produced higher[1] total work output than x^*. While simulation could be used to investigate the stochastic conditions which would make the difference between x^* and the optimal finite policy virtually indetectable, Getzler *et al.* [2] did not report the stochastic process used to derive their conclusions. Their results were not unexpected however, given that x^* was generally in the neighborhood of the optimal policy for the 320 test problems for both models.

A SEARCH PROCEDURE FOR OPTIMAL FINITE HORIZON POLICIES

Gentzler *et al.* [2] indicated that the optimal policy with a finite time horizon has the property of evenly spaced rest periods, as previously defined, and suggested that dynamic programming [1] could be used to determine optimal policies. However, the fact that x^* was generally in the neighborhood of the optimal policy for a finite time horizon suggested the following alternative approach to optimal policy determination:

(i) Compute x^* using the appropriate formula developed by Gentzler *et al.* [2]
(ii) Compute an initial estimate of the optimal number of rest breaks ($\hat{k}$) such that[2]

$$\hat{k} = \left[\frac{T - x^*}{x^* + \theta}\right]$$

[1] This is true as long as T does not correspond to an integer number of rest periods which permits a complete work period of length x^* at both the beginning and end of the time horizon. In this case, the optimal infinite horizon policy and the optimal finite policy are identical and hence, produce the same work output over the finite time horizon.

[2] $[x]$ denotes the integer portion of x.

TABLE 1. SEARCH RESULTS FOR THE LINEAR MODEL

Condition	Percent of 320 problems	Cumulative percentage
$k_0 = \hat{k}$	46.6	46.6
$k_0 = \hat{k} + 1$	53.4	100.0

TABLE 2. SEARCH RESULTS FOR THE EXPONENTIAL MODEL

Condition	Percent of 320 problems	Cumulative percentage
$k_0 = \hat{k}$	29.7	29.7
$k_0 = \hat{k} + 1$	66.6	96.3
$k_0 = \hat{k} + 2$	3.1	99.4
$k_0 = \hat{k} + 3$	0.6	100.0

(iii) Defining k_0 as the optimal number of rest periods for the finite horizon problem, compute total work output (TW) at $k = \hat{k}, \hat{k} + 1, \hat{k} + 2, \ldots, k_0, k_0 + 1$ using one of the following formulas

$$TW_l = (k + 1)(bt - at^2/2)$$

$$TW_e = (k + 1)(1 - e^{-at})/a$$

where:

TW_l = total work output during the time horizon with k evenly spaced rest periods for the linear model,

TW_e = total work output during the time horizon with k evenly spaced rest periods for the exponential model,

and $t = (T - k\theta)/(k + 1)$ = the length of the work period.

Given the assumptions of the two models, although some integer number of rest periods will be optimal, the total work output 'curve' with different evenly spaced rest periods is unimodal. Thus, k_0 is identified by the fact that $TW(k_0) > TW(k_0 + 1)$. *Theorem* 1 establishes that there is no requirement to search the TW 'curve' for $k < \hat{k}$.

Theorem 1. The optimal number of rest periods for the finite horizon problem is always greater than or equal to $\hat{k}$.

Proof. Define $T_m = (\hat{k} + 1)x^* + \hat{k}\theta$ where x^* is the optimal length of the work periods in the infinite horizon problem. If $T = T_m$ were true, then the finite time horizon would permit exactly $\hat{k}$ rest periods with complete work periods of length x^* at the beginning and end of the time horizon and $k_0 = \hat{k}$. Since $T \geq T_m$ is always true, and because k_0 is an increasing function of T, then $k_0 \geq \hat{k}$ must also be true.

COMPUTATIONAL RESULTS

The results of using this computational procedure on the 320 previously described test problems are summarized in Table 1 and 2. The results indicated that 100.0% and 96.3% of all test problems required either two or three TW computations for the linear and exponential models respectively. The maximum number of TW calculations required by the search process for any of the test problems was four. Thus, given the fairly broad set of parameters investigated, it appears that dynamic programming would not be likely to provide enough computational savings to justify the additional complexity.

REFERENCES

1. BELLMAN R (1957) *Dynamic Programming*, Princeton University Press, Princeton, New Jersey, USA.
2. GENTZLER GL, KHALIL TM, & SIVAZLIAN BD (1977) Quantitative models for optimal rest period scheduling. *Omega* **5**(2), 215–220.

Stephen E Bechtold

(*March, 1979*)

College of Business
Florida State University
Tallahassee
Florida 32306
USA

Vol. 7
1979

Capital Rationing Decision with Internal and External Funds

INTRODUCTION

LORIE AND Savage [1] were the first to use a mathematical approach in obtaining an optimum solution to a capital rationing problem. Later, Weingartner [4] proposed to formulate the problem into a mixed integer linear program under conditions of certainty. Subsequently, numerous researchers have debated on the underlying concepts of Weingartner's model and suggested alternative formulations under both certainty and uncertainty. The central issues of these debates are (1) the meaning and determination of a discount rate used in computing present values, and (2) whether the discount rate properly measures the firm's opportunity cost of capital (see, for example, Weingartner [6] for a summary of this work and of the researchers).

Earlier, Lutz [2, p. 25] distinguished two investment horizons; one is the 'fund horizon' used for available funds, and the other the 'internal investment horizon' used for internal projects. Weingartner's early model [4, p. 168] considered the availability of external funds to supplement internal funds in financing capital projects. However, most existing capital rationing studies assume a single planning horizon for alternative projects and available funds, and the availability of funds only from internal sources. In reality, projects requiring capital facilities may have different planning horizons and be financed with external funds.

In this study, we formulate the problem of rationing funds from internal and external sources to independent internal projects, assuming different planning horizons for the funds and projects. To resolve the difficulty of comparing projects with different lives, we adopt the approach suggested by Preinreich [3] in which a single project is represented by a permanent chain of cyclic investments in the project. An optimum solution maximizing the sum of the present values of net revenues from selected projects is obtained by formulating the problem into a mixed integer-linear program. The mathematical model thus formulated is applied to numerical examples.

FORMULATION

We analyze the problem of allocating available internal and external funds to independent internal capital projects from two sides, the funding side and the opportunity side.

On the funding side, an 'arrangement' represents a unique combination of a rate of interest, a maximum lending amount, and a maximum lending period. There are m arrangements of external funds, numbered 1 through m, and one arrangement of internal funds, numbered $m + 1$. Let R_i, M_i, and L_i represent the annual rate of interest, maximum lending amount, and maximum lending period for arrangement i ($i = 1, \ldots, m$), respectively. For arrangement $m + 1$, Q, instead of R_{m+1}, represents the expected rate of return on internal funds; M_{m+1} is the sum of available internal funds; and L_{m+1} is not used.

On the opportunity side, there are n independent internal projects numbered $1, \ldots, n$. We let V_{jt}, W_j, and H_j represent the expected net revenue in the tth year, initial capital requirement, and investment period for project j ($j = 1, \ldots, n$), respectively.

As to the method of paying back borrowed external funds, we assume that the principle is paid back at the end of the project's life while the interest is paid at the end of each year. The annuity of a single unit of funds from arrangement i invested in one cycle of project j is given by the following a_{ij}:

$$a_{ij} = \frac{Q(1+Q)^{H_j}}{(1+Q)^{H_j}-1}\left\{1 - R_i \sum_{t=1}^{H_j} \frac{1}{(1+Q)^t} - \frac{1}{(1+Q)^{H_j}}\right\} \quad \begin{matrix} i = 1, \ldots, m+1 \\ j = 1, \ldots, n \end{matrix}$$

which may be reduced to:

$$a_{ij} = Q - R_i. \tag{1}$$

Similarly, we obtain the annuity of the net revenue from a single cycle of project j as follows:

$$b_j = \frac{Q(1+Q)^{H_j}}{(1+Q)^{H_j}-1} \sum_{t=1}^{H_j} \frac{V_{jt}}{(1+Q)^t} \quad j = 1, \ldots, n. \tag{2}$$

We now introduce decision variables X_{ij} and Y_j representing the amount of funds from arrangement i invested in project j, whether or not project j is selected for investment respectively. Thus, the objective of the problem is to find the values of X_{ij} and Y_j that maximize the sum of the present values of annuities a_{ij} in (1) and b_j in (2) for an infinite period:

$$\text{max. } C = \sum_{i=1}^{m+1} \sum_{j=1}^{n} \frac{a_{ij}}{Q} X_{ij} + \sum_{j=1}^{n} \frac{b_j}{Q} Y_j \tag{3}$$

subject to

$$X_{ij} \geqslant 0 \qquad i = 1, \ldots, m+1; j = 1, \ldots, n \tag{4}$$

$$Y_j = 1 \text{ or } 0 \quad j = 1, \ldots, n. \tag{5}$$

If all projects are independent, the relationship (5) provides a sufficient condition. If some of the projects are mutually exclusive or dependent on other projects, additional conditions are required along the lines suggested by Weingartner [5, p. 498].

Next, we formulate the constraints on the relationships between variables. First, the initial capital investment in a project must be equal to the sum of funds invested in the project:

$$W_j Y_j - \sum_{i=1}^{m+1} X_{ij} = 0 \quad j = 1, \ldots, n. \tag{6}$$

Second, in regard to each arrangement, the sum of funds invested in the selected projects should not exceed the maximum available amount:

$$\sum_{j=1}^{n} X_{ij} - M_i \leqslant 0 \quad i = 1, \ldots, m+1. \tag{7}$$

Finally, the investment period of a project should

TABLE 1. DETAILS OF INTERNAL PROJECTS AND FUNDING ARRANGEMENTS

Details of internal projects				Details of funding arrangements			
Project number j	Required capital W_j ($1000)	Uniform annual revenue b_j ($1000)	Investment period H_j (year)	Arrangement number i	Maximum lending amount M_i ($1000)	Maximum lending period L_i (year)	Interest rate R_i
1	10	1.50	2	1	30	3	0.09
2	28	8.68	5	2	40	4	0.13
3	32	8.00	3	3	20	5	0.14
4	44	13.20	4	4	30	4	0.11
5	41	11.48	4	5	40	6	0.12
6	47	18.80	6	6*	100	6	0.15
7	50	17.00	6				
8	60	15.68	4				
9	68	16.32	3				
10	70	12.60	2				

*Internal arrangement.

TABLE 2. OPTIMAL SOLUTIONS TO THE EXAMPLE PROBLEM

Project	Problem 1 funding	Problem 2 funding
1	—	—
2	—	17 (3), 11 (6)
3	32 (6)	30 (1), 2 (4)
4	—	—
5	—	40 (2), 1 (3)
6	—	8 (5), 39 (6)
7	—	50 (6)
8	—	28 (4), 32 (5)
9	68 (6)	—
Present value	162	565

(Where $i(j)$ means i thousands of dollars raised on arrangement j).

not exceed the maximum lending period of an external arrangement from which funds are obtained:

$$(L_i - H_j)X_{ij} \geq 0 \quad \begin{matrix} i = 1, \ldots, m \\ j = 1, \ldots, n \end{matrix} \qquad (8)$$

With the objective function in (3) and the constraints in (4)–(8), the formulation of a mixed integer-linear program has been completed.

EXAMPLES

Six arrangements including an internal arrangement are considered for funding ten independent internal projects. Table 1 lists the details of the projects and arrangements. The 6th arrangement represents the internal sources of funds for which the firm expects a rate of return of 15%.

With the above data, two problems are considered. The first problem assumes the availability of only the internal funds of $100,000 while the second problem assumes the availability of both the internal and external funds; optimal solutions are displayed in Table 2. In problem 2, only arrangement 3 has not been fully used having a remaining amount of $2,000.

REFERENCES

1. LORIE JH & SAVAGE LJ (1955) Three problems in rationing capital. *Jl Busin.*, **28**, 229–239.
2. LUTZ F & LUTZ V (1951) *The Theory of Investment of the Firm.* Princeton University Press, Princeton, USA.
3. PREINREICH GAD (1940) The economic life of industrial equipment. *Econometrica*, **8**(12) 12–44.
4. WEINGARTNER HM (1963) *Mathematical Programming and the Analysis of Capital Budgeting Problems.* Prentice-Hall, Englewood Cliffs, New Jersey, USA.
5. WEINGARTNER HM (1966) Capital budgeting of interrelated projects: survey and synthesis. *Mgmt Sci.* **12**(7), 485–516.
6. WEINGARTNER HM (1977) Capital rationing: authors in search of a plot. *Jl Fin.* **32**(5), 1403–1431.

Hirohide Hinomoto

Department of Business Administration
University of Illinois
Urbana–Champaign
Illinois 61801
USA

(*July, 1978*)

Vol. 7
1979

Optimum Location of Training Centers[1]

THERE HAS been a great deal of interest in facilities location problems recently in the literature. The application developed here is based upon the set covering or p-median problems of [3] and [4]. Recent work on this problem has included both theory and applications. A specialized optimal solution procedure appears in [2]. A multiperiod generalization with a dynamic programming solution is present in [5]. Application to location of bank branches for minimization of float is made in [1]. The present application is a case study of another area of applicability of facilities location analysis: the location of training centers.

Profit and non-profit making organizations periodically conduct training sessions in one or more locations to provide initial skill training, to improve skills, to upgrade existing skills, to conduct refresher courses, or to give employees and employers an opportunity to exchange concepts and techniques. These sessions are often conducted by one or more experts in the field or by individuals from the parent company. Many of these organizations have employees located in widely dispersed areas, causing substantial incidental and travel expenses as a result of centralized training. For these training sessions which do not involve large costs for transportation or acquiring specialized equipment or training aids, rather than establishing one training center it may be less costly to conduct sessions at several training locations, thereby reducing total travel costs. However, travel cost savings will be offset by the fixed cost of establishing multiple facilities, hiring duplicate training coordinators, etc.

THE CASE

The scientific products division of a large midwestern (United States) medical supply company recognized the need to periodically retrain their instrument service technicians in advanced technology relating to the health sciences and to train them in the maintenance and upkeep of additions to their product line. Employees were located in seventeen different cities. Figure 1 shows the number and location of these employees. In order to reduce costs, the company considered the possibility of establishing one or more centers located in regions of high employee density. The cost of establishing a center and conducting a program was estimated to be approximately $1500 exclusive of room and meal charges which were believed to be constant regardless of location. It was felt, however, that in order to encourage meaningful interaction between employees and the program coordinator, a minimum of 20 participants would be required at any center. Airline travel costs for sending employees from their home city to each of the potential center locations are given in Table 1. A decision was made to locate the centers in a city which had an established company office and therefore, only those cities are considered as potential center sites.

[1] An earlier version was presented at the May 1978 TIMS/ORSA meeting in New York.

ANALYSIS

The training center problem is similar to the set covering problem of [3]. The two problems are similar in that there are fixed costs associated with opening each training center. The total costs to be minimized are the fixed costs of operating a center plus the travel costs of participants attending the training sessions. As in the fixed charge aspect of the set covering problem, the analysis of training center location problems involves a determination of which cities to establish as centers in order to minimize the total of fixed costs plus travel costs.

The problem can be formulated as an integer programming problem as follows: given (n) communities having (A_i) participants and travel costs (C_{ij}) of travelling from city (i) to city (j) per participant, determine the number of centers (m) and specify the center to which each participant should be assigned. Fixed costs at location j will be F_j if location j is utilized, zero otherwise. Denote the bivalent integer variable of the problem as X_{ij}. This will equal unity if community i assigns to center j, zero if not.

The problem is stated mathematically as:

$$\text{minimize } Z = \sum_{j=1}^{n} \sum_{i=1}^{n} A_i C_{ij} X_{ij} + \sum_{j=1}^{n} F_j X_{jj} \qquad i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, n \quad (1)$$

subject to:

$$\sum_{j=1}^{n} X_{ij} = 1 \qquad i = 1, 2, \ldots, n \quad (2)$$

$$X_{jj} \geq X_{ij} \qquad i \neq j \quad (3)$$

$$\sum_{i=1}^{n} A_i X_{ij} \geq K \qquad j = 1, 2, \ldots, n \quad (4)$$

$$X_{ij} \in \{0, 1\} \quad (5)$$

The objective function (1) and constraints (2) and (3) are necessary for the set covering structure of [3]. Constraint (4) was developed for the model as the restriction that any operating center must include at least K training participants. K is set at 20 for this case because the program coordinator needs at least 20 participants in order to achieve the desired level of meaningful interaction.

The optimum solution to this formulation was found using both integer programming and linear programming computer programs. The linear programming routine resulted in an optimal integer solution. The optimum locations are shown in Fig. 2. The minimum cost solution subject to the imposed restrictions is to establish centers at cities New York, Chicago and Los Angeles at a total cost of $12,007. The number of participants at each center will be: New York 40, Chicago 69 and Los Angeles 26. A sensitivity analysis was conducted to determine the sensitiveness of the fixed cost of establishing a training center and variability in employee density at each location.

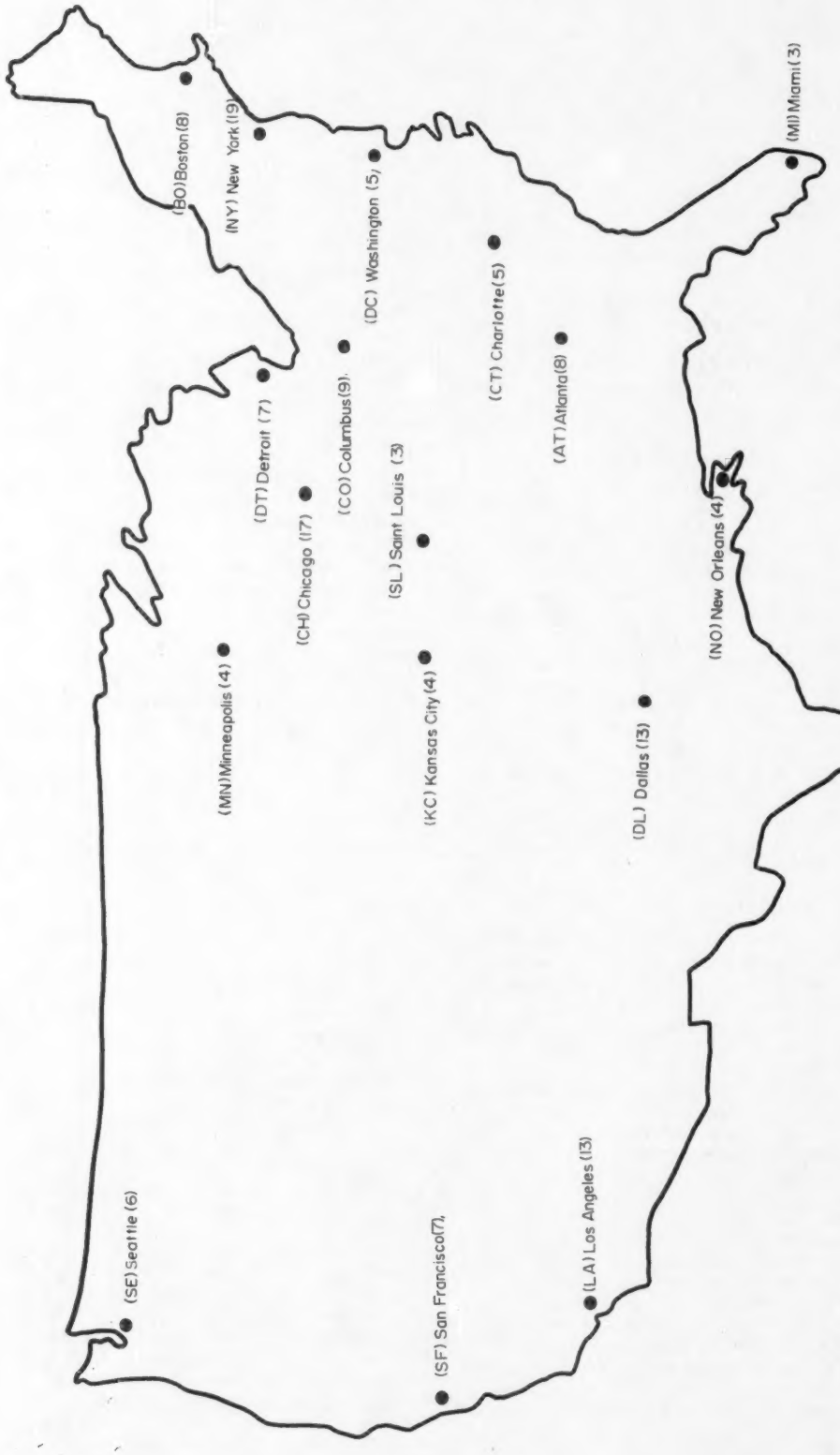


FIG. 1. *Location of existing facilities (number of employees in parentheses)*

Vol. 7
1979

Vol. 7
1979

Table 1. Total travel cost between cities for all employees (dollars)

To \ From	NY	DC	BO	AT	MI	CT	CH	MN	CO	DT	SL	DL	KC	NO	SF	LA	SE
NY	0	260	384	1008	510	490	2074	648	792	644	426	2652	680	720	2352	4368	2016
DC	988	0	624	800	450	360	1768	600	612	532	366	2340	608	624	2226	4134	1908
BO	912	390	0	1216	570	620	2380	688	1008	756	492	2938	760	808	2436	4524	2088
AT	2394	500	1216	0	312	270	1768	584	756	742	276	1586	488	328	2058	3536	1800
MI	3230	750	1520	832	0	560	3094	880	1404	1232	504	2236	760	464	2408	4108	2196
CT	1862	360	992	432	336	0	1768	768	648	658	306	1976	576	456	2198	3796	2136
CH	2318	520	1120	832	546	520	0	288	558	378	168	1716	312	552	1764	3276	1512
MN	3078	750	1376	1168	660	960	1224	0	972	686	252	1820	312	656	1554	2886	1332
CO	1672	340	896	672	468	360	1054	432	0	322	240	1950	432	544	2044	3588	1884
DT	1748	380	864	848	528	470	918	392	414	0	252	2054	448	600	1946	3614	1668
SL	2698	610	1312	736	504	510	952	336	720	588	0	1300	216	424	1736	2964	1476
DL	3876	900	1808	976	516	760	2244	560	1350	1106	300	0	336	328	1302	2080	1452
KC	3230	760	1520	976	570	720	1326	312	972	784	162	1092	0	464	1540	2626	1320
NO	3420	780	1616	656	348	570	2346	656	1224	1050	318	1066	464	0	1890	3120	1728
SF	6384	1590	2784	2352	1032	1570	4284	888	2628	1946	744	2418	880	1080	0	429	696
LA	6384	1590	2784	2176	948	1460	4284	888	2484	1946	684	2080	808	960	231	0	828
SE	6384	1590	2784	2400	1098	1780	4284	888	2826	1946	738	3146	880	1152	812	1794	0

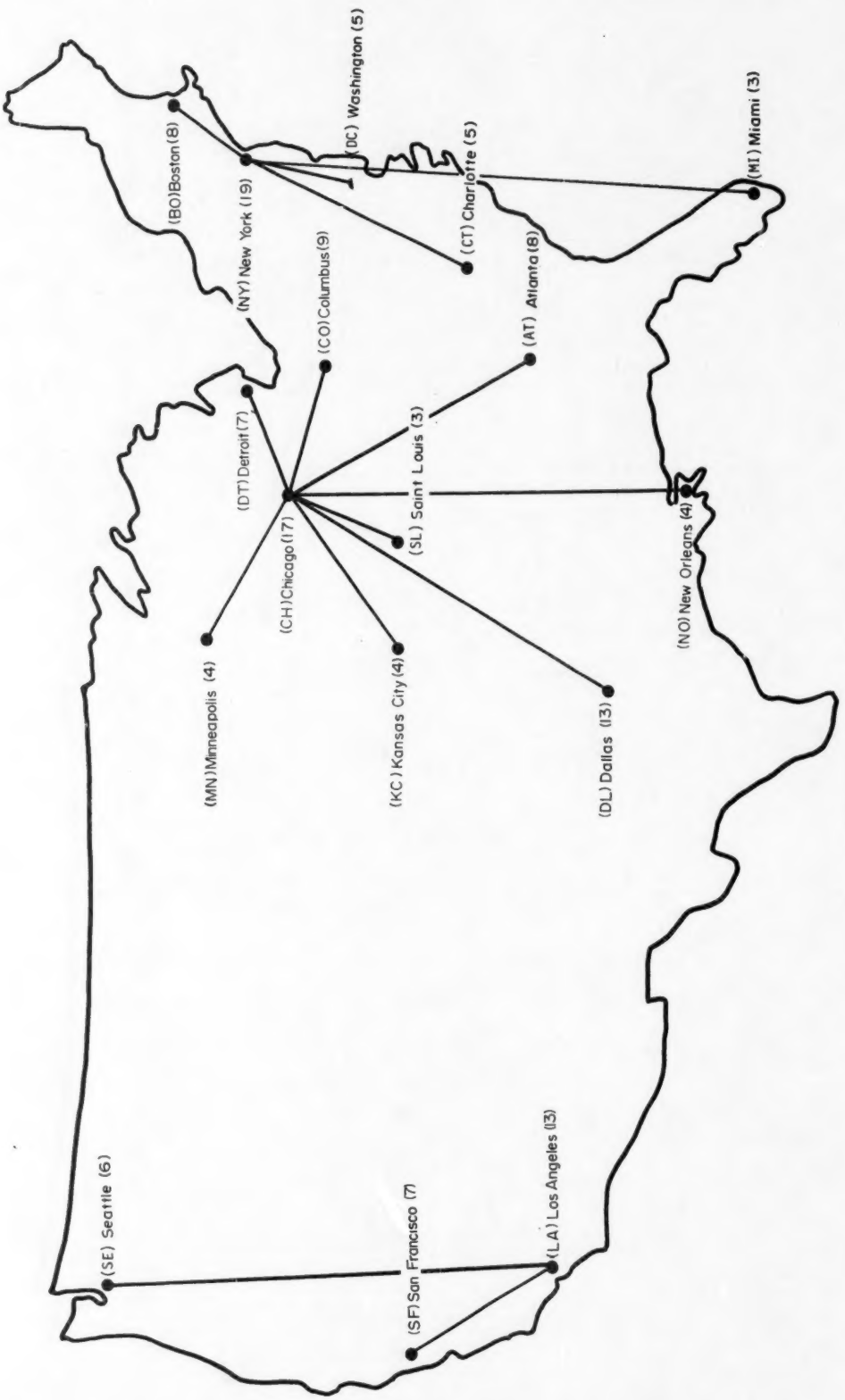


FIG. 2. *Optimal training center locations and employee assignments*

Vol. 7
1979

DISCUSSION AND CONCLUSIONS

Problems of central training center locations which are more specialized than the one presented in this article may be solved using an LP procedure. For example, a restriction on the maximum number allowed at one location may be imposed. However, the restriction will open the possibility of non-integer solutions that, while mathematically optimal, may not be realistic. A fractional optimal solution may indicate that fractions of several populations should each attend different central facilities, while a more desirable solution may be to split one or two populations only to satisfy restrictions. In any event, a linear programming solution to population allocation problems allows a decision maker the opportunity to minimize the costs of travel and facilities location. Non-integer problems, although more costly, can also readily be solved using available computer software packages.

In many applications the restriction (4) on the minimum number of participants admitted to a center may not be necessary. For the medical supply company analysis it was required that 20 employees participate in any training session in order to enhance the meaningfulness of the program to the participants. If this restriction were not imposed, the optimal solution would be to establish four centers, the additional center being located in Dallas and providing training for the Dallas and New Orleans service employees. All other allocations would remain the same.

Also in many applications, the number of training centers may be specified beforehand. In these cases an additional restriction would be required:

$$\sum_{i=1}^{n} X_{ii} = m. \tag{6}$$

This restriction requires the establishment of exactly m training centers and would be especially useful for firms which have existing facilities and are interested in minimizing total travel costs to these facilities.

REFERENCES

1. Cornuejols G, Fisher ML & Nemhauser GL (1977) Location of bank accounts to optimize float: an analytic study of exact and appropriate algorithms. *Mgmt Sci.* **23,** 789–810.
2. Narula SC, Ogbu UI & Samuelson HM (1977) An algorithm for the *p*-median problem. *Opl Res.* **25,** 709–713.
3. Revells CS & Swain RW (1970) Central facilities location. *Geog. Anal.* **2,** 30–42.
4. Torejas C, Swain RW & Bergman L (1971) The location of emergency service facilities. *Opl Res.* **19,** 1363–1373.
5. Wesolowsky GO & Truscott WG (1975) The multiperiod location–allocation problem with relocation facilities. *Mgmt Sci.* **22,** 57–65.

Larry E Shirland
Peter M Ellis

(*May*, 1979)

Department of Business Administration
Votey Building
University of Vermont
Burlington
Vermont 05405
USA

Creativity and Operational Research

CREATIVITY has been defined as the "ability to produce new forms in art or mechanics or to solve new problems" [2], and its relevance to Operational Research (defined by Wagner [13] as "a scientific approach to problem solving") is obvious. A useful starting point in considering the topic is a comment by Popper [10, p. 48], which is worth quoting in full, and to which we shall return.

> 'But it seems to me that what is essential to "creative" or "inventive" thinking is a combination of intense interest in some problem (and thus a readiness to try again and again) with highly critical thinking; with a readiness to attack even those presuppositions which for less critical thought determine the limits of the range from which trials (conjectures) are selected; with an imaginative freedom that allows us to see so far unsuspected sources of error; possible prejudices in need of critical examination.
> (It is my opinion that most investigations into the psychology of creative thought are pretty barren—or else more logical than psychological. For critical thought, or error elimination, can be better characterised in logical terms than in psychological terms)'.

The competitive importance of scientific and technological advance has led to much effort by psychologists to measure and promote creativity. Whilst this 'barren' approach was under way, much more illuminating work was, ironically, already on record in other areas of psychology, under the titles of 'learning' and 'problem-solving'. The purpose of this paper is to give a brief account of this work and its implications for OR (Operational Research). For contrast and completeness however, the 'barren' approach is mentioned first.

The first attempts at measuring creativity were based on conventional psychological testing, and led to the notions of convergent and divergent thinking. Intelligence tests require convergent thinking, and contain questions like: *Which is the odd one out*; *cow pig sheep sparrow*? 'Creativity' tests required divergent thinking, and contained such questions as: *Write down as many uses as you can think of for a brick*. High scores on the divergent thinking tests were expected to indicate creativity. Hudson [7] has carried out a vast amount of research in this area. He has shown that IQ is irrelevant to creativity. He has shown that convergence and divergence are also irrelevant to creativity, and he links creativity with "persistence, self-confidence and predatoriness".

Of the many creativity dogmas, brainstorming and lateral thinking are perhaps the best known. Neither need be expounded here. Brainstorming may be likened to an exercise in collective divergent thought, and lateral thinking has some similarities, though there is more to it than that. The suspension of criticism which is a key ingredient in brainstorming, would certainly not endear the procedure to Popper.

As indicated earlier, there are other areas of psychological investigation which are much more illuminating. This account draws together work from the fields of learning, thinking and problem solving, to form a more plausible story.

Let us start with learning, e.g. with rats in mazes, and conditioning. This, the most basic form of learning, is the process whereby a relationship between events (stimuli) results in a change in behaviour (response). An intriguing stochastic model of conditioning has been built by Walter [14, Ch. 7], who also constructed an electro mechanical representation of the model which he called "*Machina Docilis*". This behaved very much like an animal in learning experiments, and displayed other typical animal behaviour (including neuroticism).

At this rather low level of learning there is a good understanding of, for example, learning to press the red lever not the blue one to obtain food. Harlow and Harlow [5] have built on this with a long series of experiments which indicate that learning is hierarchical. They assert that learning is learned, and that items of simple learning are aggregated in learning 'sets' which are applied to further learning or problem-solving. On a higher plane, there is work on concept formation. Heidbreder [6] is generally quoted).

Thus, when we apply ourselves to a problem, we apply a complex hierarchical structure of pre-formed blocks of understanding. Much effort has gone into understanding the building of these blocks (as well as establishing their existence). However, for our purposes, their destruction is more important, because they can behave as restrictions of thought as well as tools.

Scheerer's work in this field is instructive [12]. He has carried out a series of human problem-solving experiments in which the apparent inappropriateness of earlier learning experience and the importance of habitual perception of critical objects were varied. He found three stumbling blocks to the solution of problems: assumption, habit and motivation. He describes the condition of 'stuckness' that they generate as 'fixation', because it usually manifests itself as repeated attempts at the wrong approach to the problem. Subjects made false assumptions about every aspect of a problem and failed to question those assumptions. Habitual patterns of thought and perception also hindered solution. The key to a problem might be visible and present, yet it would go unrecognised because of perceptual habit (e.g. in a problem requiring the use of string for solution, over 50% of subjects failed to 'see' string as such when it was holding up a calendar or mirror in front of them). Finally, Scheerer preesents evidence that unduly high motivation blocks the path to solution by making the subject unwilling to accept any detour that might delay achievement, and he indicates that the difference between trying personally to solve a problem, and watching someone else try, is enough in some cases to demonstrate this effect.

At this point it is as well to look back at Popper's comment, when it will be seen that the breaking of obstructive habit, assumption and previous learning is a matter only of method and logic. Here too is the irony: Hudson's book, *Contrary Imaginations*, was published in 1966, Scheerer's article in 1963. If those psychologists working on creativity had seen the connec-

Vol. 7
1979

tion between creativity and problem solving, they would have realised that they already possessed the key to progress in their field. Habit prevented them from seeing that key for what it was—fixation indeed!

Let us now turn to Edward De Bono [1] whose ideas of vertical and lateral thinking might be considered 'fringe psychology', but have relevance here. Although it is easier to grasp these ideas through examples, they can be briefly characterised. Vertical thinking is high probability thinking, strong on developing ideas, but rigidly directional; it is logic in control of mind. Lateral thinking is low probability thinking, creative, flexible; it is logic at the service of mind. De Bono explains the term thus: "It is not possible to dig a hole in a different place by digging the same hole deeper ... if the hole is in the wrong place, no amount of improvement is going to put it in the right place ... Vertical thinking is digging the same hole deeper; lateral thinking is trying again elsewhere".

This makes lateral thinking sound very much like the 'uncontrolled idea generation' of divergent thinking and brainstorming. De Bono goes on, however, to treat at length and in an abstract manner, the perception of the components of a problem, and makes it clear that habits of perception (learning sets) can obstruct the perceptual insights necessary for problem solution and for the creation of new ideas. (This is in line with Scheerer's work). He also discusses a number of techniques for deliberately breaking down obstructive habits of perception and thought. (See also Rickards [11] for a practical summary).

Looking back, we see that De Bono, Scheerer and Popper are all barking up the same tree. All are saying, in different ways, that creative or problem-solving thought depends on diligent unearthing, questioning and dismantling of assumptions, and of habits of thought and perception. Popper's suggestion that this process should be characterised in logical terms leads to the proposition that the process should be logically and systematically controlled, which is really only another way of saying that a methodology is needed.

OR has nearly as many methodologies as it has general text books, but in the light of this analysis it can be seen that they are really different attempts at the same methodology. Indeed, other management services disciplines also have methodologies which may be reduced to expressions of the same principle. For example, one might look at work study [3], where method study has a well defined methodology: Select, Record, Examine, Develop, Install, Maintain and where 'Examine' boils down to applying each interrogative (what, where, when, who, how, why) iteratively to every aspect of the situation. It is reasonable therefore to propose that there is really one true underlying methodology which consists of systematic hyper-critical examination and error elimination.

This is obviously central to the nature of OR, and although it demands considerable emphasis in the teaching of OR, it is essentially a return to first principles. The implications do, however, go further.

Whilst the purpose of OR is to dig new holes, OR is itself a hole. The pioneers of OR started a new hole, and struck ore. Their successors have made it a much bigger, deeper hole. Look again at work study; the same metaphor holds. It is simply an older, deeper hole. To use such a metaphor is of course, in no way to deny that the ore was worth getting or that very powerful and valuable varieties of shovel and other equipment have been developed.

De Bono has some pertinent observations on the properties of holes: It is unreasonable and counter to human nature, he observes, to expect people to abandon established holes without knowing where to start a new one. "The shovels of logic lie idle. There is no achievement.... No one is paid to sit around being capable of achievement...." Thus holes tend to be self-perpetuating. Further, "an expert is an expert because he understands the present hole better than anyone else except a fellow expert ... experts are usually to be found at the bottom of the deepest holes...."

Now, future practitioners of OR are likely to enter the field through the education system, and the education system is largely designed as a conducted tour of existing holes, in which knowledge is received from experts. There is a danger, therefore, that the education system will fail to transmit the heart of the OR hole, namely the art of digging new holes. This danger certainly does not go unrecognised in educational establishments, but there is a risk that the attractions of the existing hole might obscure it.

It is therefore essential to the future potency of OR that a full and explicit understanding of the nature of creativity is maintained and emphatically propagated both in the teaching of OR and amongst its practitioners.

REFERENCES

1. De Bono E (1971) *The Use of Lateral Thinking.* Pelican, Harmondsworth, UK.
2. Chaplin JP (1968) *Dictionary of Psychology.* Dell Publishing Company, New York, USA.
3. Currie RM (1968) *Work Study.* Pitman Paperbacks, London, UK.
4. Gregory SA (1972) *Creativity and Innovation in Engineering.* Butterworth, London, UK.
5. Harlow HF & Harlow MK (1949) Learning to think. *Scientific American, offprint no.* **415**, WH Freeman & Co, San Francisco, USA.
6. Heidbreder E (1946) The attainment of concepts: I, II. *Jg gen. Psychol.* **35**, 173–223.
7. Hudson L (1967) *Contrary Imaginations.* Pelican, Harmondsworth, UK.
8. Mintzberg H *et al.* (1976) The structure of 'unstructured' decision processes. *Admin. Sci. Q.* **21**, 246–274.
9. Parnes S (1971) Creativity: developing human potential. *Jl Creative Behav.* **5**, 19–36.
10. Popper K (1976) *Unended Quest.* Fontana, Glasgow, Scotland.
11. Rickards T. (1974) *Problem Solving Through Creative Analysis.* Gower Press, Epping, UK.
12. Scheerer M (1963) Problem solving. *Scientific American, offprint no.* **476.** WH Freeman & Co, San Francisco, USA.
13. Wagner HM (1969) *Principles of Operational Research.* Prentice-Hall, Englewood Cliffs, New Jersey, USA.
14. Walter WG (1961) *The Living Brain.* Pelican, Harmondsworth, UK.

RN Woolley

(*April* 1979)

University of Aston
Management Centre
158 *Corporation Street*
Birmingham B4 6TE
UK

BLANK
PAGE

Vol. 7
1979

OMEGA. The Int. Jl of Mgmt Sci., Vol. 7, No. 6, pp. 577 to 580
© Pergamon Press Ltd 1979. Printed in Great Britain
0305-0483/79/1101-0577$02.00/0

International Register of Theses and Dissertations

The following are obtainable at the Interlibrary Loans Section, JB Priestly Library, The University of Bradford, Bradford 7, West Yorkshire, UK:

ALLEN J (1967) The Motivation of Mail-Order Agents. MSc Thesis, Management Centre, University of Bradford.

BALLANGER NJ (1973) UK Financial Institutions and the Management of UK Industry. MSc Thesis, Management Centre, University of Bradford.

BARNETT AB (1973) A Systems Dynamics Model of an Oilfields Development. PhD Thesis, Management Centre, University of Bradford.

BARROW PJ (1977) The Changes in Information in Published Accounts, 1946–1976. MBA Thesis, Management Centre, University of Bradford.

BEVERIDGE CW (1973) Manpower Planning and its Significance for the Company. MSc Thesis, Management Centre, University of Bradford.

BIDWELL RO (1974) The Impact of Stricter Pollution Controls on Product Costs. PhD Thesis, Management Centre, University of Bradford.

BONNINGTON PJ (1971) The Car Industry and the Common Market. MSc Thesis, Management Centre, University of Bradford.

BRADWELL A (1977) The Provision of Financial Information to Employees. MBA Thesis, Management Centre, University of Bradford.

BROWN D (1970) Business Budgeting. MSc Thesis, Management Centre, University of Bradford.

BUDDIN AD (1977) Sales Forecasting in a British Group of Companies. MBA Thesis, Management Centre, University of Bradford.

BURKE MCP (1973) An Assembly Line Balance for the Tele-Communication Industry—the 2000 Type Selector Machines. MSc Thesis, Management Centre, University of Bradford.

CHAN FK (1976) The Asian Dollar Market. MBA Thesis, Management Centre, University of Bradford.

CHOW SL (1977) A Survey of Operational Research Models to Evaluate Advertising Effectiveness. MBA Thesis, Management Centre, University of Bradford.

CLARKSON AH (1973) Strategic Planning in the Aerospace Industry. MSc Thesis, Management Centre, University of Bradford.

CLAYBOURN BH (1977) A Review of Dispatching Rules: Comparison of Various Dispatching Rules with Real Data Drawn from several different Manufacturing Industries. MBA Thesis, Management Centre, University of Bradford.

COLLEDGE NR (1977) Legal, Decent, Honest, Truthful—*Caveat Emptor, Caveat Vendor*: a Study and Review of *Advertising Practice with Regard to Societal, Substantiation and Control Issues*. MBA Thesis, Management Centre, University of Bradford.

COLLINS JB (1971) The Organizational and Educational Status of the Procurement Function: a Study in Selected Areas of British Industry. MSc Thesis, Management Centre, University of Bradford.

CONNORS MFJ (1972) An Examination of Management by Objectives as a Technique Enhancing the Motivation of Managers. MSc Thesis, Management Centre, University of Bradford.

COWELL DW (1973) The Effectiveness of Marketing Courses. PhD Thesis, Management Centre, University of Bradford.

CREW MA (1971) The Theory of Peak Load Pricing and its application. PhD Thesis, Management Centre, University of Bradford.

CURRY JF (1977) Strategic Planning in the Commercial Banking Industry. MSc Thesis, Management Centre, University of Bradford.

DAVIES AP (1969) Financial Information, Company Law Reform and the Stock Market as a Resource Allocator. MSc Thesis, Management Centre, University of Bradford.

DOWNING MJ (1973) Taxation of Capital Transfer: Capital Transfer Tax in the UK and a Comparison with the Death and Gift Duties of other Members of the EEC. MBA Thesis, Management Centre, University of Bradford.

DRURY JC (1972) An Information System for Production Control. MSc Thesis, Management Centre, University of Bradford.

DURBER PW (1976) The Evolution of the City Code on Takeovers and Mergers. MSc Thesis, Management Centre, University of Bradford.

ELLSMORE MA (1974) Evaluating Logistics Service—a Model Building Approach. MSc Thesis, Management Centre, University of Bradford.

ENDERWICK P (1976) An Examination of Wage Leadership Models in the Post-War Period. MBA Thesis, Management Centre, University of Bradford.

FARRER PJ (1972) Productivity Bargaining—Feasibility Study MSc Thesis, Management Centre, University of Bradford.

FORD D (1971) The Marketing of Central Heating with Particular Reference to the Channels of Distribution. MSc Thesis, Management Centre, University of Bradford.

FRASER JC (1978) The Role of Marketing in the Dominant Firm Situation. MSc Thesis, Management Centre, University of Bradford.

GIBB WF (1969) Interfirm Comparison—an Appraisal. MSc Thesis, Management Centre, University of Bradford.

GIBSON BRC (1973) An Appraisal of Group Technology and its Implications for a Company in the Machine Tool Industry. MSc Thesis, Management Centre, University of Bradford.

GUHA PS (1973) The Significance of Cultural Anthropology in Marketing Strategies. MSc Thesis, Management Centre, University of Bradford.

HAMMOND A (1971) Job Shop Forward Loading. MSc Thesis, Management Centre, University of Bradford.

HANRAHAN JM (1970) A Study of Attitudes Towards an Effectiveness of Direct Mail Advertising to Medical Practitioners in the United Kingdom. MSc Thesis, Management Centre, University of Bradford.

HARDWICK MR (1977) An Examination of the Practical Value of Quantitative Methods for Pricing Decisions. MBA Thesis, Management Centre, University of Bradford.

HARDY AG (1972) An Investigation into some Areas of Industrial Relations Existing in Engineering. MSc Thesis, Management Centre, University of Bradford.

HARKNESS BL (1971) Establishment of an Effective Inventory Control System for a Company in the Machine Tool Industry. MSc Thesis, Management Centre, University of Bradford.

HART NA (1977) Advertising Education: a Research into the Value of Professional Qualifications in the Advertising Industry. MSc Thesis, Management Centre, University of Bradford.

HAWKINS K (1970) The Reform of Collective Bargaining in British Industry. MSc Thesis, Management Centre, University of Bradford.

HERBERT PFA (1972) Underwriting Issues of Corporate Securities. MSc Thesis, Management Centre, University of Bradford.

HILL AB (1969) The Measurement of Work Variety. PhD Thesis, Management Centre, University of Bradford.

HINDLE MW (1970) An Investigation of Public Relations in the Pharmaceutical Industry. MSc Thesis, Management Centre, University of Bradford.

HOLGATE PA (1976) Value Added Tax and the Retailer. MSc Thesis, Management Centre, University of Bradford.

HOLMES PW (1969) Supervisor's Training in the Engineering Industry: a Case Study. MSc Thesis, Management Centre, University of Bradford.

HOOLEY P (1973) The New Issues Market: an Examination of the Methods of Marketing a New Equity Issue. MSc Thesis, Management Centre, University of Bradford.

HOPWOOD NA (1975) Some Productivity Determinates of a Semi-Automatic Machine Centre in Engineering Batch Production. MSc Thesis, Management Centre, University of Bradford.

HOWARD K (1977) A Computer Modelling System for Resource Planning in Local Government. PhD Thesis, Management Centre, University of Bradford.

MARQUES JP (1977) Foreign Investment Decision: the Theory and a Managerial Response in Brazil. MBA Thesis, Management Centre, University of Bradford.

MILLER PD (1977) The Effect of Uncertainty on Scheduling. MBA Thesis, Management Centre, University of Bradford.

MORGAN DJ (1977) Corporate Social Objectives and Social Value Testing. MBA Thesis, Management Centre, University of Bradford.

MORGAN JPW (1976) The Development of Merchant Banking within the UK Financial System, with Particular Reference to 1960–1973. MSc Thesis, Management Centre, University of Bradford.

MORRISON DJ (1977) The Application of the Marketing Concept to Retail Pharmacies. MSc Thesis, Management Centre, University of Bradford.

NAWAZ T (1978) Insurance Legislation and Portfolio Management Policies. MBA Thesis. Management Centre, University of Bradford.

NEWTON DJ (1974) The Impact of US Foreign Direct Investment in the United Kingdom. MSc Thesis, Management Centre, University of Bradford.

NG WLK (1977) The Training of Supervisors: a Reassessment. MBA Thesis, Management Centre, University of Bradford.

NORRIS PW (1977) A Study of the Development of a Staff Appraisal System in a Large Organisation. MSc Thesis, Management Centre, University of Bradford.

NOTTINGHAM C (1974) Survey of Computer Related Fraud. MSc Thesis, Management Centre, University of Bradford.

OLDMAN AM (1975) Standardization of Flexibility in Financial Reporting. MSc Thesis, Management Centre, University of Bradford.

OR CH (1976) An Empirical Study of a Systems Dynamics Cash Management Model. MBA Thesis, Management Centre, University of Bradford.

POWELL TR (1975) Marketing Cost Analysis. MSc Thesis, Management Centre, University of Bradford.

QUAYLE NJT (1971) Controlling a Company's Transport Operations. MSc Thesis, Management Centre, University of Bradford.

SMITH BD (1977) Financing Problems of Small Growth Companies. MBA Thesis, Management Centre, University of Bradford.

SOUTHALL M (1975) Employee Participation: its Meaning, Main Forms and Desirability, MSc Thesis, Management Centre, University of Bradford.

SPEHT A (1973) A Financial Analysis of Technology Applied to Large Scale Welfare Catering. MSc Thesis, Management Centre, University of Bradford.

STAINFORTH JAL (1974) Cost Models in Urban Transport Planning. MSc Thesis, Management Centre, University of Bradford.

SUGDEN JP (1974) The Uses and Applications of Annual Budget: an Empirical Case Study of Budgetary Control. MSc Thesis, Management Centre, University of Bradford.

SYKES EA (1972) Credit and the Independent Retailer. MSc Thesis, Management Centre, University of Bradford.

TAN CN The Marketing of Tourism. MBA Thesis, Management Centre, University of Bradford.

TAYLOR PR (1977) Some Economic Aspects of Eating Hunted Fish and the Prospect for a Fish Farming Industry. MSc Thesis, Management Centre, University of Bradford.

TEO CL (1976) New Product Policy for Corporate Survival and Growth. MSc Thesis, Management Centre, University of Bradford.

THORNTON N (1974) Financial Information Classification for Management. PhD Thesis, Management Centre, University of Bradford.

TIEN SING YOUNG PJ (1977) Development Banking in Mauritius. MBA Thesis, Management Centre, University of Bradford.

TOWNSEND TG (1978) The Characteristics of Corporate Death: an Investigation into the Use of Predictive Models in Predicting Failure or Acquisition. MBA Thesis, Management Centre, University of Bradford.

TREMLETT TR (1974) The Euro-Currency and Capital Markets. MSc Thesis, Management Centre, University of Bradford.

TUOMEY R (1974) Capital Project Investment. MSc Thesis, Management Centre, University of Bradford.

TURPIN IJ (1977) Currency Management in the Multi-National Company. MSc Thesis, Management Centre, University of Bradford.

URREA JD (1977) Evaluation and Development of the Family Firm. MBA Thesis, Management Centre, University of Bradford.

VINCZE JW (1976) Ethnic Background and its Effect on Consumer Behaviour. PhD Thesis, Management Centre, University of Bradford.

VONG JHC (1977) A Survey of Mathematical Models in Financial Theory. MBA Thesis, Management Centre, University of Bradford.

WHELAN RW (1975) Trade Union Responses to the Growth of the Multinational Enterprise. MSc Thesis, Management Centre, University of Bradford.

WILMOTT RA (1974) The Measurement of Efficiency: Return on Capital or Added Value. MSc Thesis, Management Centre, University of Bradford.

WOODBRIDGE P (1974) The Importance of Businessmens' Motivations and Values in the Analysis of Corporate Objectives. MSc Thesis, Management Centre, University of Bradford.

YOUNG PM (1977) Risk Management in Local Government. MBA Thesis, Management Centre, University of Bradford.

The Following are obtainable at the Main Library, University of Guelph, Guelph, Ontario, Canada N1G 2W1:

ANANDAJAYASEKERAM P (1978) Economics of a Milk Transportation Firm. PhD Thesis, School of Agricultural Economics, University of Guelph.

BOUMA GG (1977) Identification and Evaluation of Factors Facilitating or Impeding Limited Resource Farmers Receptive to Change in Establishing Farm and Family Viability. MSc Thesis, School of Agricultural Economics, University of Guelph.

BROOKSBANK RD (1976) Agricultural Land Resources and Land Use in an Urbanizing Region. MSc Thesis, School of Agricultural Economics, University of Guelph.

EOONOUS MAA (1976) An Application of a Multi-Attribute Attitude Model to Consumers' Selection of Supermarkets. MSc Thesis, School of Agricultural Economics, University of Guelph.

Hall DJ (1978) A Model of Coordination for the Counselling of Adolescents on the Community Level. MSc Thesis, School of Extension Education, University of Guelph.

Keats CW (1978) Assessment of the Need for Pre-Retirement Education. MSc Thesis, School of Extension Education, University of Guelph.

Kolodkine M (1976) An Application of Multiperiod Linear Programming to Total Operation Planning of a Fertilizer Bulk Blender Firm. MSc Thesis, School of Agricultural Economics, University of Guelph.

Leal GR (1976) Structural Development in Latin America. MSc Thesis, School of Agricultural Economics, University of Guelph.

Macaulay TG (1976) A Recursive Spatial Equilibrium Model of the North American Beef Industry for Policy Analysis. PhD Thesis, School of Agricultural Economics, University of Guelph.

Maicher GB (1976) An Analysis of Economic Opportunities for Farm Adjustments on Selected Dairy Farms Shipping Farm-Separated Cream and Industrial Milk in Cans. MSc Thesis, School of Agricultural Economics, University of Guelph.

Percival S (1977) Teacher Preparation for Overseas Work: a Case Study. MSc Thesis, School of Extension Economics, University of Guelph.

Seguin RE (1976) An Economic Evaluation and Interpretation of the Countryside as a Natural Resource. MSc Thesis, School of Agricultural Economics, University of Guelph.

Swegles C (1977) A Study of the Staff Training Program at the University of Guelph. MSc Thesis, School of Extension Education, University of Guelph.

Trant MJ (1976) Classification of Limited Resource Farmers in Ontario Based on Behavioural and Economic Characteristics. MSc Thesis, School of Agricultural Economics, University of Guelph.

Yu GH (1976) A Benefit–Cost Study of Herbicide Use in Corn Production in Ontario. MSc Thesis, School of Agricultural Engineering, University of Guelph.

Zwart A (1977) An Empirical Analysis of Alternative Stabilization Policies for the World Wheat Sector. PhD Thesis, School of Agricultural Economics, University of Guelph.

Information for inclusion in this Section, presented in the appropriate form, should be sent to the Editor of Omega.

Vol. 7
1979

OMEGA The Int Jl of Mgmt Sci. Vol. 7, No. 6, p. i
Pergamon Press Ltd 1979. Printed in Great Britain

ERRATA

Omega, Volume 7, Number 4, 1979

The bottom line of Table 1, p. 323, should read:

UNSCEAR [25]

On page 323 line 27 in the second column should read:

... by others [1, 9, 12, 23, 25, 26] gives an average risk ...